Market Structure and College Access in the US

Emily E. Cook*

February 5, 2024

Abstract

In response to rising college tuition, state governments have implemented subsidies in the form of grant aid to students. Understanding the effect of these subsidies implemented at scale requires a equilibrium model in which differentiated, competing colleges set tuition, financial aid policies, and admission standards. I develop such a model with careful attention to the geographic nature of the market. I show demand-side and equilibrium responses to an illustrative counterfactual policy: an additional $1,000 in grant aid for Louisiana students to attend an in-state public college. While student responses alone would increase enrollment by 5.7%, Louisiana public colleges respond by simultaneously increasing tuition levels and price discrimination, partially offsetting the demand-side effect of the policy to generate a net enrollment increase of 3.4%.

JEL Codes: I23, I28, L13, L31
Keywords: college choice, college admission, college enrollment, financial aid, mixed oligopoly, non-profit firms, demand estimation

*Department of Economics, Tulane University, ecook4@tulane.edu. I thank Sarah Turner, Federico Ciliberto, and Gaurab Aryal for their invaluable advice on my dissertation, which laid the foundation for this paper. I also thank Flavio Cunha, Michael Dinerstein, Ian Fillmore, Jeremy Fox, Chao Fu, Doug Harris, Jean-Francois Houde, Adam Kapor, Ekaterina Khmelnitskaya, Lance Lochner, Chris Neilson, Jeff Smith, Kevin Stange, Todd Stinebrickner, Chris Taber, and Basit Zafar for helpful conversations. This work benefited from input from participants at many conferences and seminars. Ed Hall and the Advanced Research Computing Services at the University of Virginia, and Brian Streicher and Technology Services at Tulane University provided computational resources and technical support. Victoria Roloff provided excellent research assistance. The Bankard Fund for Political Economy at UVa and the National Science Foundation provided financial support. The results contained herein were derived in part from data provided by ACT Research, and in part from data provided by the College Board. Copyright © 2000-2015 The College Board. www.collegeboard.org. All errors are mine.

1 Introduction

Inflation-adjusted tuition and fees have risen 125% at public four-year colleges in the last thirty years, and 80% for private four-year colleges during the same period (Ma and Pender 2022), leading to concerns about college access and affordability.[1] Because college completion is associated with higher earnings and increased economic mobility,[2] policymakers seek to improve college affordability through policies such as government grant aid to students (Page and Scott-Clayton 2016; Herbaut and Geven 2020). In equilibrium, these policies affect student decisions *and* the tuition, aid policies, and admissions standards set by colleges. As famously suggested by Bennett (1987), colleges may respond to the availability of financial aid by increasing tuition, diminishing the effect of the policy.

To understand how public and private colleges' tuition, financial aid policies, admission standards, and student enrollment respond to subsidies, I develop an equilibrium model in which public and private universities compete strategically for students. I build on previous work on on student application and enrollment choices (Arcidiacono 2005; Howell 2010; Kapor 2020; Bleemer 2021), as well as previous work on equilibrium in higher education (Epple, Romano, and Sieg 2006; Fu 2014; Epple et al. 2017; Epple et al. 2019). The most significant differentiator between this work and other equilibrium models is that I model the incentives facing *individual* colleges, accounting for the geographic nature of the market. This approach allows me to simulate responses to regional policies, such as the types of policies typically implemented at the state level. This framework also allows me to examine policy responses across individual institutions, to understand how public institutions actions affect private colleges, and to understand to what extent state-level policies affect colleges in neighboring states.[3]

My equilibrium model features four stages: first, colleges choose tuition, aid policies, and an admission threshold; second, high-school seniors observe colleges' choices and exogenous characteristics and choose a set of colleges to apply to; third, colleges offer admission and (potentially) financial aid to a subset of applicants; and finally, students make their enrollment decisions. Students base their application decisions on the expected value of an application portfolio, taking into account their probability of admission and like-

[1] Increases in aid have partially offset these tuition increases (Cook and Turner 2022).

[2] See Heckman, Lochner, and Todd (2006), Goldin and Katz (2008), and Chetty et al. (2020).

[3] Other work has extended Epple, Romano, and Sieg (2006) in different directions, for example Gordon and Hedlund (2019a) and Gordon and Hedlund (2019b) imbed monopolistically competitive colleges into an incomplete markets, life-cycle model to study the causes of the increase in tuition since 1987. A different approach to modeling price discrimination is presented in Fillmore (2023). He models aid as the result of an auction in which very selective private colleges bid for students.

lihood of receiving financial aid. A multiple-discrete choice problem arises in the application stage because a student can choose more than one school to apply to at a time. I model the application stage as a choice between application portfolios, which converts the multiple-discrete choice problem to a single-discrete choice problem.[4] At the enrollment stage, students choose the single college that provides the highest utility from among colleges that admitted the student. On the supply-side, colleges have heterogeneous preferences over profit and other factors, including the admission rate, the academic fit of the student body, and the mean income of the student body.

The empirical method incorporates the regional nature of competition in this market. In general, students must travel away from home to attend a four-year college, making distance a key determinant of college choice. As a result, competition between colleges is regional. Some colleges, such as Harvard, draw students from all over the nation, while some colleges (particularly less-selective public institutions) draw the vast majority of their students from in-state. I capture this feature of the market empirically using data on enrollment from each student state of residence to each college from the Residence and Migration Survey of the Integrated Postsecondary Education Data System (IPEDS).[5] For this empirical market definition, I use *historic* data relative to my period of study in order to avoid defining markets based directly off of the model outcome. I estimate the model using data on applications and enrollment by college and student state of residence. More precisely, I use ACT and College Board to proxy applications.[6] To estimate features of the admissions process, college financial aid functions, and some aspects of heterogeneity across students, I supplement the student state-by-college aggregate data sources with moments computed from student survey data, specifically, the National Center for Education Statistics (NCES)' High School Longitudinal Study of 2009 (HSLS:09).

I estimate the model in several steps. First, I obtain the federal and state grant aid functions and admissions parameters from regressions conducted on the student survey data. Then I estimate demand and supply separately, extending traditional approaches to estimation of differentiated-products models of oligopoly competition, as in Berry (1994) and Berry, Levinsohn, and Pakes (1995), to this setting. Notable differences

[4]This solution has been used in other contexts. For example, Gentzkow (2007) estimates demand for newspapers, where consumers can choose to read more than one newspaper. In his model, consumers choose between "bundles" of the available newspapers. See his paper for a more thorough review of multiple-discrete choice models.

[5]IPEDS is a set of surveys of the population of colleges in the United States and is maintained by the National Center for Education Statistics.

[6]Students send score reports to the colleges that they apply to. I have the number of score reports sent by high school graduating seniors in each state to each college, which I use to proxy applications. To the extent that students may sent some score reports without following through with a complete application, I may underestimate application costs.

between the higher education market and a typical consumer product market motivate substantial deviations from the standard model. The institutional details that drive these differences include: 1) colleges are not profit maximizing, 2) consumers (students) must apply and be accepted in order to "purchase" the product, and 3) colleges engage in third-degree price discrimination—they allow prices to vary by student admission test scores and family income. Thus, I supplement the traditional market share measures with aggregate and individual moments that help identify parameters of the application and admissions process.

To illustrate properties of the demand-side model, I run several counterfactual simulations holding the college tuition, aid policies, and admissions parameters fixed. Dropping price at all colleges across the nation by 1% would increase enrollment by 1.07%, all else equal. The enrollment effects are strongest at middle income levels; for example, students with family incomes between $48 and $75k would see a 1.21% change in enrollment from a 1% decrease in all college tuition. Decreasing admissions standards by 1% is estimated to increase total enrollment by 0.63%, with the strongest effects at the lowest income levels.

Using the estimated model, I simulate pricing and enrollment outcomes after changes in the generosity of a statewide financial aid program. As a specific example to illustrate mechanisms in the model, I implement a $1,000 grant to students in Louisiana to attend in-state public universities. The demand-side effect alone, holding the supply-side fixed, increases the total number of students enrolled to 1,015 students, or 5.7% of all Louisiana enrollment. When public universities are allowed to respond by changing prices and tuition levels, they increase tuition levels and merit-based financial aid — in other words, they increase the degree of price discrimination. After accounting for these responses by public universities, the enrollment increase is 3.4%. The effect of the subsidy combined with the public university response would decrease enrollment at private universities competing for Louisiana students by up to 3%, with the largest effects concentrated on less-selective private universities located in or near Louisiana. Forthcoming work will add additional dimensions of heterogeneity in the utility function, expand the counterfactuals to allow private universities to respond to the policy, and allow universities to change admissions standards in the counterfactual.

2 Data

Estimation of the model relies on data on college characteristics and student choices at both the aggregate and individual levels. These data sources are described in subsections 2.1 and 2.2. The model is informed by an empirical description of market structure, which is the subject of section 2.3.

2.1 College Data

Table 1 summarizes the college-level information, which comes from IPEDS for the academic year 2013-14.[7] The main source of price variation is between the in-state public tuition at a mean of $8,370, and the out-of-state public and private tuition levels, which are $19,850 and $32,470 on average, respectively. Student grant aid from all sources (federal, state, and institutional) can make up a large share of tuition — or even exceed tuition at some colleges. For example, an average student with a family income between $30,000 and $48,000 would receive $10,070 in total grant aid at a public college, generating a net tuition of $8,370-$10,070 = -$2,330 at an in-state public college. The excess grant aid can typically be used to help pay for room and board.

Selectivity measures include the 25th percentile ACT score within the freshman entering class, and college rankings from Barron's Profiles of American Colleges. ACT scores are measured on a scale from 1 to 36, and the mean among test-takers was approximately 21 in 2013. Public colleges tends to be less selective than private colleges; the average of the 25th percentile ACT score is 20.1 at public colleges compared to 22.3 at private colleges. Barron's rankings provide another (discrete) measure of selectivity. The private institutions studied here are much more likely than public institutions to be listed as "Very Competitive" or higher by Barron's.

I include several other college characteristics as controls: the degree completion rate (measured as completion within 6 years of enrollment), whether the college has a graduate program, whether the college has a religious affiliation, whether the college is a historically black college or university (HBCU), discrete measures of the campus size, and the urbanicity of the location of the campus. I also include the Carnegie Classification, which largely a function of the number and type of degree programs offered at the institution. The last panel of Table 1 summarizes these characteristics.

[7] IPEDS is a collection of college surveys on a range of topics such as admissions and enrollment, finances, financial aid, and degrees granted. These surveys include the entire population of post-secondary institutions in the US that participate in federal student aid programs under Title IV (such as Pell Grants and federal student loans). I focus on four-year public or private non-profit institutions located in the 48 contiguous states or the District of Columbia. I include two-year colleges and for-profit institutions in the outside option along with the labor market, because the substitution between the for-profit four-year and traditional public and private non-profit four-year sectors is limited. For-profit institutions comprised approximately 3 percent of first-time four-year enrollment among recent high school graduates in the 2013-14 academic year. After imposing some restrictions on size and selectivity, the population for analysis includes 421 public colleges and 454 private colleges. I restrict analysis to institutions that are classified in the Carnegie Classification system as Baccalaureate, Masters, or Doctoral/Research level colleges. Finally, I restrict the sample to institutions that are somewhat large or selective. Institutions included have either: total undergraduate enrollment greater than 2,000, a Barron's selectivity rank of "Competitive" or higher, or receive applications from more than 2% of students in at least one state.

2.2 Student Data

To estimate the college choice model, I use aggregate data on applications and enrollment by student state of residence, as well as microdata on student decisions from the NCES. The IPEDS Residence and Migration Survey provides enrollment in each institution by student state. I use these data to generate enrollment shares by state, defined as a college's enrollment from a state as a proportion of the state's high school graduating class. The second column of Appendix Table B2 shows the distribution of enrollment shares across all states and institutions. This distribution is heavily right-skewed, with the enrollment share reaching only 1% of high school graduates at the 90th percentile, and then 17% at the maximum.

I use data from the SAT and the ACT admissions tests as a measure of applications by student state of residence. When students take the SAT or ACT, they have the option to send their score reports to institutions to supplement their college applications. Because the colleges in my study require an SAT or ACT score report to accompany each application, these reports provides a proxy for applications.[8] The College Board and the ACT each provided a table showing the total score reports sent to each institution in the country, by student state of residence and graduation year. I link the data from the ACT and SAT by college name and then sum the total score reports across the two tests.[9] An example of these data for Connecticut is presented in Table B1. I construct an "application penetration" measure using the score-send data by dividing the total number of score reports sent to each institution from each state by the number of high school graduates from that state. The application penetration measure is summarized in the first column of Table B2. The maximum value is 0.38, which means that at most, a university receives applications from 38% of high school graduates in a single state. The average application penetration is 0.02, meaning that on average across college-state pairs in my sample, 2% of students from the state send an application to the college.

In addition to these sources of aggregate data, I have access to a sample of high school students from the National Center for Education Statistics' High School Longitudinal Study of 2009 (HSLS:09). This survey provides information on the application portfolios, admission outcomes, enrollment decisions, and student

[8] Card and Krueger (2005) show a high correlation between test-score sending and applications, so the SAT and ACT report-sending data may be a close approximation for applications. However, to the extent that sending a score report is easier than sending an actual application, my estimates may overstate the number of applications sent and understate admissions rates. This should affect interpretation of the application numbers in my results, but it does not affect the enrollment numbers and characteristics of enrollees, which are of primary importance for estimating the supply-side. Further, since I measure selectivity using characteristics of enrollees, this should not affect interpretation of universities' admission standards.

[9] This does imply some double-counting in the case that a student sends *both* an ACT and an SAT score with their application, which is likely an unusual situation. To the extent this double-counting exists, it will be reflected in a lower application cost in my estimates.

characteristics for a sample of students who were freshmen in high school in 2009.

2.3 Empirical Features of the Market Structure

College enrollment in the US is highly regional (Hoxby 1997; Smith, Pender, and Howell 2018; Hoxby and Turner 2019; Fu et al. 2022). In fact, 74% of four-year college enrollees attended a college in their state of residence as their first primary institution, based on student records data from the HSLS:09. Yet, competition between colleges is linked nationally. A regional college that serves only residents of Virginia may compete for Virginia students with another less regional college, such as Duke, which in turn competes with a regional college in Georgia for Georgia residents. In this way, the regional colleges in Virginia and in Georgia compete indirectly. This is similar to the chain-linked competition among local newspapers studied by Fan (2013).

A careful analysis of strategic competition in this market incorporates the regional nature of enrollment while recognizing that the market is integrated through chain-linked competition. To do this, I develop a method based on Fan (2013) to generate sets of competitor institutions in each state. Mechanically, the method is as follows:

1. Generate enrollment shares for each college, j, and state of residence, l: $e_{jl} = \frac{n_{jl}}{N_l}$. The numerator is the number of high school graduates from state l who attended college j, and the denominator is the number of high school graduates from state l.[10]

2. Sort the data by college and descending enrollment share.

3. Starting from the top of the list for each college, select the states that together account for at least 85% of the college's total enrollment.[11][12]

In these computations I used the eight years of data *prior* to the year of this analysis (2005-06 through 2012-13) to smooth out year-to-year fluctuations and to avoid using current-year enrollment patterns to define the sets of competitors in each market.

As an example, Figure 2 demonstrates the outcome of this method for Tulane University, Loyola New Orleans, Louisiana State University (LSU) and Louisiana Tech. Tulane University is a well-known private

[10] The total number of high school graduates by state is obtained from NCES data as collected in the *Knocking at the College Door* report produced by the Western Interstate Commission for Higher Education (WICHE).

[11] A larger enrollment threshold would capture more of a college's demand, while a smaller threshold limits the computational challenges that arise from larger choice sets.

[12] I also require that at least 0.15% of students in the state apply to the college.

university, and its enrollment comes from a large number of states. Still, Tulane draws more enrollment from southern states than a similar institution further north would (e.g. Georgetown University). Loyola New Orleans is a private Catholic university with a more local draw of enrollment. LSU is the flagship public university, so it enrolls students primarily from Louisiana, but also attracts a non-negligible share of students from Texas. Finally, Louisiana Tech draws the overwhelming majority of students from Louisiana, as is the case for many regional public universities. The example makes clear that this method—which is based on enrollment shares by state—does not automatically choose the largest states in terms of population. This is why, for example, Tulane is marked as competing in Vermont but not in North Carolina.

A drawback of this method is that counterfactual conclusions are conditional on this definition of regional competition: there is no role for expanding the regional draw of enrollment, even if a college were to become more attractive by reducing prices or admission standards. This is a reasonable assumption if the regional draw of enrollment is primarily a function of the knowledge students have about their options — e.g. the reason a student in New York would not choose a regional college in Louisiana is in part because the student is not aware of the college in Louisiana. Further, even if there is a margin for counterfactual pricing and aid policies to broaden the regional draw of enrollment, the changes would have to be very large to impact the conclusions substantially. The average enrollment share in the state-university pairs dropped by my regional market definition is 0.0001.[13]

3 Model

There are two types of agents in the model: students, i, and colleges, j.[14] There are $J+1$ colleges $j = 0, 1, ..., J$, where $j = 0$ represents the outside option, including for-profit colleges, open-admission or non-selective four-year colleges, two-year colleges, and the labor market. The timing of the model is illustrated in Figure 1. First, each college simultaneously chooses tuition, a financial aid policy, and an admission policy. Each student then observes the institutions' choices and exogenous characteristics and chooses a set of colleges to apply to—an *application portfolio*. After applications are received, colleges notify students of their admission outcomes. Each student then chooses which college, if any, to enroll in from among the colleges that have offered admission.

[13] This is after filtering to four-year institutions located in the US and with Baccalaureate, Masters, or Doctoral Carnegie Classification codes, but without imposing further restrictions on size as described in Section 2.1.

[14] Students are high school graduating seniors and colleges include most four-year colleges in the US. See Appendix Section **??** for details.

The following sections cover details about functional forms and the information available to each agent at each stage. In the first two subsections below, I describe the financial aid and admissions policies. Next I describe the student's problem at the enrollment and application stages, and then move to the supply-side.

3.1 Pricing

Before making an enrollment decision, student i will observe the *net tuition* that they would pay at each college, defined as tuition minus grants received from the federal and state governments and the college. The net tuition student i pays to attend college j is given by:

$$p_j(inc_i, s_i, l_i, \chi_{ij}) = t_j(l_i) - a^f(inc_i) - a^s(inc_i, s_i, l_i) - a^c_j(inc_i, s_i, \chi_{ij}), \tag{1}$$

where $t_j(l_i)$ is tuition at college j, which varies across states only by in-state status for public colleges.[15] $a^f(inc_i)$ is the level of the Federal Pell Grant as a function of income inc_i. State grant aid $a^s(inc_i, s_i, l_i)$ is a function of the student income level and observed college admission test score, s_i, as well as features of the state grant aid program available in the student's state, l_i. Finally, $a^c_j(inc_i, s_i, \chi_{ij})$ is college j's grant aid to student i as a function of the student's income, test score, and an idiosyncratic determinant of aid, χ_{ij}, which follows the standard normal distribution. Students do not observe χ_{ij}, but they observe the aid offer if the college accepts the student. The aid offer appears simultaneously with the admission notification, for the purpose of this model. Thus to notate the price from the student's perspective, I use $p_j(inc_i, s_i, l_i, a^c_{ij})$.

The government aid functions are deterministic and are estimated outside of the model, as discussed in Section 5.1. The college aid function merits further discussion here, because the parameters of this aid function are endogenous choices of the college. Colleges choose aid levels as a function of student income and test scores as well as some information about the student that is idiosyncratic (χ_{ij}). The specification allows for the existence of merit-based and need-based aid depending on the degree to which the college penalized income and rewards test scores in the aid function. In general, students are quite uncertain about their financial aid outcomes before applying, an empirical reality that is captured by the idiosyncratic component in the aid function. As it is infeasible to model aid as continuous, I discretize college aid into four levels. The value that determines the aid level is the aid index $\tilde{a}^c_{ij} = a^c_{j1} + a^c_{j2}s_i + a^c_{j3}inc_i + \chi_{ij}$, where the

[15] Reciprocity agreements give some out-of-state students in-state status for the purpose of pricing. I will take these agreements as given in the analysis.

coefficients vary by college j. The aid index maps to an aid amount according to the function below:

$$a_j^c(inc_i, s_i, \chi_i) = \begin{cases} 0 & \text{if} \quad \tilde{a}_{ij}^c \leq k_1 \\ \$1,280 & \text{if} \quad k_1 < \tilde{a}_{ij}^c \leq k_2 \\ \$5,430 & \text{if} \quad k_2 < \tilde{a}_{ij}^c \leq k_3 \\ \$21,230 & \text{if} \quad \tilde{a}_{ij}^c \geq k_3 \end{cases}. \tag{2}$$

The values are chosen to match midpoints of quartiles of the distribution of aid among enrolled students. Colleges can thus choose to what degree they consider merit (ACT scores) or income in the aid policy by adjusting the coefficients on those two factors in the aid index.

3.2 Admissions

Admissions outcomes for each student are based on an admission index, $s_{ij} = s_i + \zeta_{ij}$, where s_i is the student's college admission test score, and $\zeta_{ij} \sim N(0, \sigma_\zeta^2)$ is an idiosyncratic measure of student fit with the particular college. Prior to the admission stage (Stage 3 in Figure 1), each college j only knows the distributions of test scores s_i and the student fit ζ_{ij} among all high school graduates. At the admission stage, the values of s_i and ζ_{ij} for all of college j's applicants are revealed to college j. College j can then construct the admission index for each applicant and compare the index to the pre-determined admission threshold, $\underline{s}_j$, to determine the admission outcome.

Student i knows the value of s_i, but knows only the distribution of ζ_{ij}. The student also knows the admission threshold for each college j, and can thus construct the conditional probability of admission. Some additional notation is useful for representing the admissions probability. Let $\mathbf{D}_i$ be a $(J+1) \times 1$ vector, where $D_{ij} = 1$ if i was offered admission at college j and $D_{ij} = 0$ otherwise, whether this is because the student did not apply or because the college rejected the student. Further, let $\mathbf{Y}_i$ be a vector indicating a student's application portfolio. $\mathbf{Y}_i$ is a $(J+1) \times 1$ vector where the element $Y_{ij} = 1$ if the student applies to school j and 0 otherwise. Then the probability that a student is accepted to j conditional on application is:

$$\mathscr{P}(D_{ij} = 1 | s_i, \underline{s}_j, Y_{ij} = 1) = \mathscr{P}(s_i + \zeta_{ij} > \underline{s}_j) = \Phi\left(\frac{s_i - \underline{s}_j}{\sigma_\zeta}\right), \tag{3}$$

where $\Phi(\cdot)$ is the standard normal CDF.

3.3 Student Utility from Enrollment

With the net tuition and admission probability as specified above, the utility student i receives from enrolling at college j is:

$$u_{ijl}(inc_i, s_i, l_i, X_{ij}, a^c_{ij}) = \beta_{0i} + \beta_{1i} p_j(inc_i, s_i, l_i, a^c_{ij}) + \boldsymbol{\beta_2} X_{ij} + \xi_{jl} + \eta_{ijl}.$$

Students may always choose not to attend any college ($D_0 = 1$), in which case they receive a mean utility which is normalized to zero.

This specification allows the value of attending any college, β_{0i}, to vary by student. In particular, $\beta_{0i} = \beta_{00} + \beta_{01} inc_i$. Similarly, the disutility of net tuition varies by income level of the student, so that $\beta_{1i} = \beta_{10} + \beta_{11} inc_i$. Students may also value exogenous characteristics of the university, $\mathbf{X_{ijl}}$, including (but not limited to) the distance from the student's residence to the college, whether or not the college is in-state, the public/private status of the college, and various controls for quality. In addition, the utility function includes a measure of college quality that is observed by the student but not by me, captured in ξ_{jl}. Finally, the student-college idiosyncratic component of utility, η_{ijl}, is distributed Type I Extreme Value. Students learn this value after application (e.g., this could represent information about campus dining options presented during a campus tour).

Due to the Type I Extreme Value assumption on η_{ij}, the probability that a student enrolls in a particular college among colleges that admitted the student is a conditional logit probability. If the college does not admit the student, then the probability of enrollment is zero. Let $\mathbf{a_i^c}$ be the set of college aid offers that the student faces. Then, letting $\nu_j(inc_i, s_i, l_i, X_{ij}, a^c_{ij}) = u_{ij}(inc_i, s_i, l_i, X_{ij}, a^c_{ij}) - \eta_{ij}$, the probability that i enrolls in j given a set of admission and aid offers can be written:

$$\mathbb{P}(Enroll_{ij} = 1 | \mathbf{D}_i, inc_i, s_i, l_i, \mathbf{X_i}, \mathbf{a_i^c}) = \begin{cases} \frac{e^{\nu_j(inc_i, s_i, l_i, X_{ij}, a^c_{ij})}}{\sum_{j' \ni D_{ij'}=1} e^{\nu_{j'}(inc_i, s_i, l_i, X_{ij}, a^c_{ij})}}, & \text{if } \mathbf{D}_{ij} = 1 \\ 0, & \text{if } \mathbf{D}_{ij} = 0. \end{cases} \tag{4}$$

3.4 Application Value Function

Each student chooses an application portfolio to maximize their expected utility of enrollment net of application costs. To derive this expected value, we can first find the utility i receives from being accepted to $\mathbf{D}_i$. This is the expected value of the maximum utility among the colleges with $D_{ij} = 1$:[16]

$$v(\mathbf{D}_i|inc_i, s_i, l_i, X_i, \mathbf{a_i^c}) = ln\left(\sum_{j\ s.t.\ D_{ij}=1} e^{\nu_j(inc_i, s_i, l_i, X_{ij}, a^c_{ij})}\right) + \gamma, \tag{5}$$

where γ is the Euler–Mascheroni constant. This constant cancels out of the choice probabilities as it does not change the relative utility of any option, and will be dropped from here on.

Recall that $\mathbf{Y_i}$ is a vector representing the application set. If the student applies to j, $Y_{ij} = 1$, otherwise $Y_{ij} = 0$. The probability of admission to j conditional on application, $\mathscr{P}(D_{ij} = 1|s_i, \underline{s}_j, Y_{ij} = 1)$, is described in subsection 3.2. Because the unobserved component of the student application index is independent across colleges, the student has the following probability of being admitted to a set $\mathbf{D}_i$:

$$\begin{aligned}&\mathscr{P}(\mathbf{D}_i|s_i, \mathbf{Y}_i)\\ &= \prod_{j\ni Y_j=1,\ D_j=1} \mathscr{P}(D_{ij} = 1|s_i, \underline{s}_j, Y_{ij} = 1) \prod_{j\ni Y_j=1,\ D_j=0} [1 - \mathscr{P}(D_{ij} = 1|s_i, \underline{s}_j, Y_{ij} = 1)].\end{aligned} \tag{6}$$

Let $\mathbf{A}$ represent the set of possible aid offers from universities that admitted the student, and $P(\mathbf{a^c}_i|inc_i, s_i, \mathbf{D}_i)$ represent the probability of receiving a specific combination of aid offers. The value of applying to Y_i is the expected utility of application net of application costs:

$$\begin{aligned}&V(\mathbf{Y}_i|inc_i, s_i, l_i, \mathbf{X_i}) =\\ &\left[\sum_{\mathbf{D}_i\subseteq\mathbf{Y}_i} \mathscr{P}(\mathbf{D}_i|s_i, \mathbf{Y}_i) \sum_{\mathbf{a^c}_i\subseteq\mathbf{A}} (P(\mathbf{a^c}_i|inc_i, s_i, \mathbf{D}_i)v(\mathbf{D}_i|inc_i, s_i, l_i, \mathbf{X_i}, \mathbf{a^c}_i))\right] - c_i(\mathbf{Y}_i),\end{aligned} \tag{7}$$

where $c_i(\mathbf{Y})$ represents i's cost of application to $\mathbf{Y}$. The cost depends on the application cost associated

[16]The log-sum formula follows from the assumption that η_{ij} are Type-I Extreme Value with location and scale parameters normalized to 0 and 1, respectively.

with each college-market pair, plus an idiosyncratic component that varies across application sets:

$$c(\mathbf{Y}_i, l_i) = \left(\sum_{j \ni Y_{ij}=1} c_{jl} \right) + \epsilon_{i\mathbf{Y}_i}, \tag{8}$$

where $c_{jl} = \beta^{\mathbf{c}} X_{jl}^c + \xi_{jl}^c$. The idiosyncratic cost, $\epsilon_{\mathbf{Y_i}}$, is distributed Type I Extreme Value. Students may have different costs to apply to different sets of colleges for many reasons, including school and parent guidance.[17]

3.5 Aggregation to Enrollment Shares and Application Penetration

Based on the model so far, I can construct enrollment shares and application penetration measures, which are useful for estimation. First, note that each student faces a unique set of application portfolios from which to choose (a choice set), which I denote Υ_i. These choice sets are in part determined by the student's geographic location. The choice sets are discussed in further detail in Section 5.3.

Conditional on the choice set Υ_i and student characteristics, the probability that a student applies to an application portfolio $\mathbf{Y}$ is:

$$\mathbb{P}(Apply_{\mathbf{Y}} | inc_i, s_i, l_i, \mathbf{X_i}, \Upsilon_i) = \frac{e^{\tilde{V}(\mathbf{Y}|inc_i, s_i, l_i, \mathbf{X_i})}}{\sum_{\mathbf{Y}' \in \Upsilon_i} e^{\tilde{V}(\mathbf{Y}|inc_i, s_i, l_i, \mathbf{X_i})}}, , \tag{9}$$

where $\tilde{V}(\mathbf{Y}|inc_i, s_i, l_i, \mathbf{X_i}) = V(\mathbf{Y}|inc_i, s_i, l_i, \mathbf{X_i}) - \epsilon_{i\mathbf{Y}}$.

Aggregating across portfolios that contain j for each student and across all students yields the application penetration for college j for a given state:

$$\mathbb{P}(Apply_j | l_i = \ell) = \frac{1}{N_\ell} \sum_{i \ni l_i = \ell} \sum_{\mathbf{Y} \ni Y_j = 1} \mathbb{P}(Apply_{\mathbf{Y}} | inc_i, s_i, l_i, \mathbf{X_i}, \Upsilon_i) \tag{10}$$

In estimation, this model-derived application penetration will be matched to the observed proportion of students who apply to each college in each state. Note that the sum of the market penetration of applications yields a number greater than one, because each application portfolio will be counted as many times as colleges it contains. In other words, if a student applies to three colleges, that student is counted in the

[17] The specification does allow some probability that a student may find it less costly to apply to a large application set rather than a smaller subset of the same colleges, but the likelihood of this situation is small if c_{jl} is relatively large for each school. The assumption greatly facilitates estimation of the model.

market penetration of applications for each of the three colleges.

Another equation used in estimation is the enrollment share by student state. For each student, the probability of enrolling in college j, unconditional on admission decisions, is:

$$\mathbb{P}(Enroll_j|s_i, inc_i, l_i, \mathbf{X_i}, \Upsilon_i) = \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \mathbb{P}(Apply_{\mathbf{Y}}|s_i, inc_i, l_i, \mathbf{X_i}, \Upsilon_i) \times \sum_{\mathbf{D}\in\mathbf{Y}} \mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}_i)\mathbb{P}(Enroll_j|inc_i, l_i, \mathbf{X_i}, \mathbf{D}) \tag{11}$$

Each component in this expression (as indicated by the brackets) has been described before. They are, from left to right, the probability of application to a specific application portfolio conditional on the student's characteristics and choice set (equation 9), the probability of acceptance to a set of colleges conditional on the student test score and colleges applied to, (equation 6), and the probability of enrollment conditional on the student's admission and aid outcomes (after integrating out the aid outcomes in equation 4). Aggregating to the market-level yields the following expression for the enrollment share:

$$\mathbb{P}(Enroll_j|l_i = \ell) = \frac{1}{N_\ell} \sum_{i\ni l_i=\ell} \mathbb{P}(Enroll_j|s_i, inc_i, l_i, \mathbf{X_i}, \Upsilon_i). \tag{12}$$

Section 5.4 demonstrates how I use the enrollment shares and application penetration in estimation.

3.6 The College Objective Function

In most market analysis, firms are assumed to maximize economic profit and have no preference over consumer characteristics. But in the context of the US higher education market, the firm objective function must be adapted to account for colleges' preferences over student characteristics. Previous research on the supply-side incentives in US higher education agrees on this point, though specific functional forms vary (Rothschild and White 1995; Epple, Romano, and Sieg 2006; Fu 2014; Epple et al. 2017; Epple et al. 2019; Blair and Smetters 2021).

In this model, universities potentially care about profit (π_j), but also consider their admission rate (r_j), the mean academic fit of the student body ($\tilde{s_j}$), and a measure of mean income (μ_j^{inc}).[18] All of these objects

[18] $\tilde{s_j}$ represents the percentile rank of student i's admission index in the population of all students—those enrolled at j, enrolled elsewhere, or not enrolled. The conversion to percentiles prevents certain students from being infinitely better than others from the university's point of view (the support of ζ_{ij} is all real numbers). If the contribution of the student admission index to the college

are functions of all university choices (tuition levels, aid coefficients, and admission thresholds), but to keep notation concise I drop the dependence on the endogenous variables in the notation. Universities may vary in the degree to which they are concerned about each of these variables. Each university maximizes:

$$W_j(t_j(l_i), \mathbf{a_j^c}, \underline{s}_j) = \pi_j^{(1-\lambda_{1j}-\lambda_{2j}-\lambda_{3j})} \tilde{s_j}^{\lambda_{1j}} r_j^{\lambda_{2j}} \left(\frac{1}{\mu_j^{inc}}\right)^{\lambda_{3j}}, \tag{13}$$

where

$$\pi_j = \underbrace{\sum_{i \ni Enroll_{ij}=1} t_j(l_i) - a^c(s_i, inc_i) + \psi_j}_{\text{Revenue}} - \sum_{i \ni Enroll_{ij}=1} \underbrace{(\alpha_1 + \alpha_2 q_j + \alpha_3 X_j^s + \omega_j - ispub_{ij}(\alpha_4 + \omega_j^{is}))}_{\text{Cost}}. \tag{14}$$

The first bracket notates total revenue. This is tuition, $t_j(l_i)$— which varies by in-state status for public universities, minus the college aid amount, plus ψ_j, which represents revenue from private gifts and federal, state, and local appropriations. The second bracket represents the student's cost to the university. This varies by university characteristics and quality, but importantly also by the number of students enrolled, q_j. The number of enrolled students is derived from the demand system described above. Allowing the per-student cost to depend upon current enrollment implies a (potentially) increasing marginal cost. Finally, public colleges may have an incentive to enroll in-state students, which reduces the effective marginal cost: $ispub_{ij}(\alpha_4 + \omega_j^{is})$. The parameter α_4 will be estimated, and will represent the difference in revenue between in-state and out-of-state students.[19]

Colleges choose tuition ($t_j(l_i)$) the aid policy parameters on income and ACT scores (a_{j1}^c and a_{j2}^c) and an admission threshold ($\underline{s}_j$) to maximize Equation 13. As mentioned above, tuition levels are constrained to be the same across student locations except in the case of public universities, which may charge different tuition levels for in-state and out-of-state students. Tuition, aid policy parameters, and admissions standards are solutions to the first-order conditions of the college problem. To find the university-specific parameters, this system of first-order conditions can be solved for λ_{1j}, λ_{2j}, λ_{3j}, the marginal cost, and in the case of

objective function was unbounded, it would be possible to maximize a college's objective function by pushing the admission threshold to infinity and admitting an infinitesimally small number of students.

[19] Note that the data are unclear as to what portion of state appropriations and other revenue sources are linked to in-state students, so I must estimate the revenue difference between these student types. This is identified by pricing variation conditional on other differences in in-state and out-of-state student characteristics.

public universities, the differential cost for in-state students, $(\alpha_4 + \omega_j^{is})$.

4 Identification

In any model of supply and demand, the main identification challenge is to separate the effect of endogenous variables on supply and demand, knowing that observed values of endogenous variables are generated by both sides of the market. This affects the current cross-sectional study in two ways. First, differences across institutions in tuition, financial aid policies, and admission thresholds may be correlated with unobserved determinants of demand, in particular, the ξ_{jl} in utility. This follows because universities make their decisions about tuition, aid policies, and admissions standards taking into account expected demand, which is a function of ξ_{jl}. So, for example, I cannot assume that $E[t_j \xi_{jl}] = 0$ when estimating demand. Second, when estimating the effect of enrollment quantity on marginal cost, α_2, I cannot assume that the university's enrollment is independent of unobserved determinants of cost, e.g. I cannot assume $E[q_j \omega_j] = 0$. This follows because q_j is a function of tuition, aid policies, and admissions standards, which are in turn a function of ω_j. Thus, I must identify the effect of price and admissions standards on utility and the effect of enrollment on marginal cost without using these conditions.

4.1 Demand-Side

The instrument for net tuition in demand estimation is an indicator for in-state public institutions. Indicators for whether an institution is in-state and the private/public status of the institution are included in utility directly, and the instrument is the interaction between these two variables. Thus, the identification of the price effect relies on two differences: the difference between enrollment at public and private institutions in- and out-of state, conditional on state fixed-effects, distance to each school, and other measures of quality. Violations of this assumption would require that attending an in-state public institution is valued inherently, aside from the subsidized in-state price and other factors included in my controls, such as distance to college.[20]

[20] One might argue that loyalty to the in-state public college sports team might result in a violation of this assumption. However, in general the inclusion of certain controls for college sports activity (such as NCAA membership) does not significantly alter the estimates. It is unlikely that sports team loyalty plays an important role in college decisions at most schools, for most students.

4.2 Supply-Side

On the supply-side, the institution-specific parameters and the marginal cost are found as solutions to the first-order conditions of the college's problem given in Equation 13. After finding the value of the marginal cost for each university, I set it equal to the model expression for marginal cost ($\alpha_1+2\alpha_2 q_j+\alpha_3 X_j^s+\omega_j$). The identification challenge is to isolate the effect of enrollment (q_j) on marginal cost. Enrollment is a function of college choices, which are in turn a function of unobservable cost shifters (ω_j). I need an instrument that is uncorrelated with the cost-shifters but is correlated with enrollment. The instrument I use is enrollment predicted as a function of distance alone, using the estimated effect of distance from the demand estimation process. In constructing the instrument, I also need to set aside the role of the admission threshold, as it is a function of cost-shifters. So, I predict enrollment as function of distance alone where students have 100% probability of admission at every university. In essence, I set aside every determinant of demand except for distance when I generate the instrument.

5 Estimation and Results

Estimation proceeds in several steps: first, I estimate the federal and state aid functions and the admissions probability using individual student data. These steps are described in Sections 5.1 and 5.2. Then I simulate a representative sample of 1,000 students in each state, with students varying by location, ACT score, and family income. This step is discussed in Appendix A. After I have a simulated sample of students, I simulate each student's choice set, as described in Section 5.3. I estimate demand and college financial aid policies simultaneously using the simulated method of moments. Finally, taking the demand estimates as given, I estimate supply based on the universities' first-order conditions. Subsections 5.4 through 5.6 describe these final two steps.

5.1 Federal and State Aid Functions

The price that students pay for college is determined in part by federal and state grant aid, as shown in Equation 1. The Pell Grant is by far the federal government's largest college grant aid program, and awards aid largely on the basis of need. State grant aid amounts are generally determined by student family income or some combination of income and merit. I estimate Federal Pell and state grant aid policies in the

HSLS:09 survey using enrolled students. Throughout the rest of the estimation procedure, the parameters of these policies are fixed.

For federal aid, I run a Tobit regression of Pell grant aid amounts on family income and family income squared. Pell aid is bounded below at \$0 and above at \$5,650, which is the maximum Pell grant set by the government in the applicable year. I exclude students with incomes above \$75k, assuming that they will not receive any Pell grant.[21] The results for the federal aid equation are shown in Table 2. The predicted Pell grant at a family income of \$30k is \$3,795 and at a family income of \$60k is approximately \$822.

State grant aid is more complicated, varying across states both in total generosity and in the degree to which the state program rewards merit or need. I distinguish between three groups of states: states with large merit aid programs, California (which is large enough to run a state-specific regression and has a program that is based solely on need), and all other states, which have relatively small programs that typically tend to put more emphasis on need. The list of states with large merit aid programs is derived from Fitzpatrick and Jones (2012).[22] For each type of state, I run a Tobit regression of student state aid on student characteristics, whether the institution is public or private, and the generosity of the state's grant aid program as measured by the total state grant aid budget per high school graduate. Aid amounts are bounded below by 0. Note that student state grant eligibility is determined by the student state of residence in these regressions. I assume that the aid is only applicable at in-state institutions, which is typically the case with these grant aid programs. The estimates from these regressions are shown in Table 3 are consistent with expected patterns: for example, in states with large merit aid programs, a student with a family income of $30,000$ is expected to be offered \$467 more in state grant aid than an identical student with a \$100k family income. In states that focus less on merit aid and more on need-based aid ("All Other States" in the table), the income gradient is much steeper: a student with a family income of \$30k is expected to be offered \$3,230 more than an identical student with a family income of \$100k. A regression for California that includes the ACT score shows that the ACT does not have a economically or statistically significant effect on aid received in that state, so I leave out the ACT score measures for California.

[21] Approximately 8% of students in the HSLS:09 who have family incomes above \$75, attend a college in my sample, and have National Student Loan Database (NSLDS) records received a non-zero Pell grant. However, the overall rate of Pell receipt is even lower for this income range because students who did not receive any federal aid are not in the NSLDS system. Imputing zero Pell grant for students without NSLDS records yields a rate below 5%.

[22] States with large merit aid programs at this time were: AR, GA, FL, KY, LA, MD, MI, MS, NM, NV, SC, SD, TN, WV, and WY.

5.2 Admission Probability

The admission probability as described in Section 3.2 depends on only the university's threshold, the student's ACT score, and an idiosyncratic measure of academic fit. The assumption that the idiosyncratic component is normally distributed leads to a simple expression for the probability of admission, given in Equation 6.

In the HSLS:09 data, I observe students' application sets and admission outcomes, and could therefore estimate the parameters of Equation 6 using a Probit regression on these data, provided I observe the students' test scores and the university thresholds. A complication arises here because university thresholds are unobserved, although I do see student admission test scores. While I don't observe admission thresholds, I do have information on selectivity; in particular, I have the 25th percentile of test scores among enrolled students at each university. I assume that the threshold is a quadratic function of this measure and estimate the parameters that determine the threshold from the Probit regression on admission outcomes. To be clear, the estimating equation is a Probit where the probability of admission is given by:

$$\mathscr{P}(D_{ij} = 1 | s_i, \underline{s}_j, Y_{ij} = 1) = \mathscr{P}(s_i + \zeta_{ij} > \underline{s}_j) = \Phi\left(\frac{s_i - \underline{s}_j}{\sigma_\zeta}\right),$$

where $\Phi(\cdot)$ is the standard normal CDF. Defining $ACT25_j$ as the observed 25th percentile of the ACT score among admitted students at college j, I assume $\underline{s}_j = \beta_0^{adm} + \beta_1^{adm} ACT25_j + \beta_2^{adm} ACT25_j^2$. Note that since the coefficient on s_i is constrained to be zero, I can identify the standard deviation σ_ζ in this regression.

The estimates of σ_ζ, β_0^{adm}, β_1^{adm}, and β_2^{adm} are provided in Table 4. Figure 3 shows the admission threshold function, marking the admission threshold for certain example universities. The admission threshold can also be interpreted as the ACT score that gives a student a 50% chance of admission if idiosyncratic components of the application are average. The graph shows that the admission threshold for Vanderbilt was approximately 35, while the threshold for Tulane was approximately 25. A regional public university like Nicholls State in Louisiana was much less likely to reject students, with an admission threshold of approximately 12. Another way to think of this is that admission to Vanderbilt was far from guaranteed even at the highest possible ACT score of 36, while an average ACT score of near 20 would nearly ensure admission to Nicholls State.

5.3 Choice Sets

As discussed in Section 3, each student may apply to any school within their individual-specific choice set. In my setting, choice sets are both exogenous and unobserved, so I must make an assumption about the distribution of choice sets. There is a wide range of assumptions that would rationalize observed data. One potential assumption is that all students have full information and consider every college. At the other extreme, I could assume every student always applies to every college in their choice set, so choice sets are indistinguishable from application portfolios (in this case, there would be effectively no choice at the application stage).

All research on the application portfolio choice problem must make some assumption about the distribution of choice sets, but the assumptions vary. Howell (2010) utilizes an approach that allows estimation of a full-information model, and Arcidiacono (2005) uses a sampling approach to select choice sets from an exogenously determined distribution. I use an assumption that students' choice sets are limited by their state and by random information that generates differences in choice sets across students in the same state. This assumption is most similar to the approach used by Arcidiacono (2005). I use this assumption both because limited information is most consistent with the institutional setting, and because it helps resolve the computational difficulty of estimating the demand-side.[23]

Substantial computational savings are also generated by limiting each student to send a maximum of four applications. In the ELS:2002 data, which includes primarily students who would apply to college in 2004, 86% of college applicants apply to four or fewer of the four-year colleges that are included in my analysis, so the assumption is not terribly restrictive.[24]

In Section 2.3, I described the geographic limitations on a college's market. This corresponds to a geographic limitation on each student's choice set, as students must choose from only the colleges that compete in the student's state. After applying this limitation, the market with the largest number of colleges has just over 150 colleges competing for these students (see Appendix Figure B1). Combining this assumption with the limitation that students apply to no more than four colleges still yields nearly 21 million possible portfo-

[23]The computation is challenging because students have a portfolio choice problem: the number of possible application sets gets very large with even a small number of colleges in the choice set. In addition, for every application portfolio there are many different admission outcomes and aid offers, complicating the computation of the expected value of a given admission portfolio.

[24]The HSLS:09 data do not record all applications sent by each student, so it is not possible to get a corresponding number for students applying to college in 2013. The best information I have from that survey says that 78% of all applicants apply to four or fewer institutions among the entire population of post-secondary institutions (including 2-year institutions and other schools not included in my analysis). Of course, the share of applicants applying to four or fewer colleges in my subset of four-year colleges studied is larger than 78%.

lios in a state in which 150 colleges compete. I further limit the problem by making restrictions on the types of colleges that are combined in a single application portfolio, following a similar strategy in Kapor (2020). Evidence from the individual data suggest that it is highly unlikely that students apply to, for example, three highly selective institutions (like Harvard, Yale and Stanford) and a non-selective college. To establish these results empirically, I examine patterns of selectivity in the ELS data using Barron's rankings from the 2004 *Profiles of American Colleges*.[25] I characterize application portfolios by the number of colleges in each of six selectivity categories. I find that 95% of students' application portfolios fall in the top 74 combinations of selectivity categories, so I limit the possible portfolios to follow these patterns.

After applying these restrictions, I simulate choice sets of 20 colleges randomly from the colleges that compete in the student's state. These choice sets correspond to the Υ_i in the model. There is no data to inform the likelihood of specific choice sets, but each option must be observed at least enough times that the applicant share data can be rationalized (i.e. a college that receives applications from 25% of all high school graduates must be in the consideration set for at least 25% of students). To achieve this property when sampling, I weight each university by the application share in the corresponding state.

5.4 Estimation Strategy for College Choice and Financial Aid Policy Parameters

I use the method of moments to simultaneously estimate the enrollment utility and application cost parameters as well as parameters of the college financial aid policy, as in Equation 2. College-level financial aid policy parameters must be estimated because they are not observed in the data. This is similar to how admissions thresholds are unobserved and are therefore estimated as discussed in Section 5.2. In this section, I first describe the rationale for estimating the college financial aid policies simultaneously with the demand-side of the model. Then I discuss the moments I use in estimation and further details of the estimation process.

Federal and state aid policies can be estimated separately from the rest of the model, as described in Section 5.1, because there is no selection on unobservables.[26] College-level financial aid policies, on the other hand, cannot be estimated separately from the rest of the model due to a selection problem. College use information other than test scores and income levels to set aid amounts. Students will see their aid amount only upon admission to the college, and then make their enrollment decisions based on this aid amount.

[25] I use ELS here instead of the HSLS:09 because the HSLS:09 only records details on up to three applications per student.

[26] This assumption reflects published policies, where aid amounts are generally determined by income levels and/or test scores.

Thus, students with higher aid amounts will select into college attendance. This implies that in a sample of *enrolled* students such as my HSLS:09 sample, the aid amounts will generally be higher than expected for the average high school graduate with the same ACT score and income. By estimating the college's aid policies together with the demand system, I account for the selection process through the demand system.

While I do not observe college aid policies directly, I do observe much information that is informative about these policies. From the college-level IPEDS data, I observe the net tuition paid by enrolled students from five different income levels after all sources of aid (federal, state, and institutional) have been applied. Table 1 shows the average net tuition by income group and public/private college type from these data. In the HSLS:09 student survey, I see the identity of each enrolled college and the amount of aid each student received from the college, as well as the student's family income and ACT score.

Recall that the college aid a student receives depends upon the aid index $\tilde{a}^c_{ij} = a^c_{j1} + a^c_{j2} s_i + a^c_{j3} inc_i + \chi_{ij}$, where the a^c_{j1}, a^c_{j2}, and a^c_{j3} are the college policy parameters. These are estimated as functions of observed data:

$$\begin{aligned} a^c_{j1} &= \beta_1^{aidc} X_{1j}^{aidc} \\ a^c_{j2} &= exp(\beta_2^{aidc} X_{2j}^{aidc}) \\ a^c_{j3} &= -exp(\beta_3^{aidc} X_{3j}^{aidc}). \end{aligned} \tag{15}$$

With the specification for the financial aid policy in hand, I estimate the enrollment utility, application cost, and financial aid policy parameters using the moment of moments. The first two moment equations are based on the enrollment and application shares. The instruments for mean utility, $\mathbf{Z_{jl}}$, contain all exogenous college characteristics in the utility function as well as the instrument for price as discussed in Section 4. I assume that these instruments are uncorrelated with the unobserved component of mean utility, ξ_{jl}, and thus can use the moment $E[\mathbf{Z_{ijl}} \xi_{jl}] = 0$. Similarly, I assume that $E[\mathbf{X^c_{jl}} \xi^c_{jl}] = 0$ in the application cost function.

Estimation of coefficients on interactions between college and student characteristics requires that I match moments that capture variation across students in their preferences for a specific college. Such moments come from the student-level data. In particular, I match: the average distance to college among enrolled students and its square, and the average family income of enrolled students at public and at private colleges. The values of these moments are shown in the first panel of Table 5.

To estimate the aid policy parameters, I match regression coefficients from an auxiliary model estimated

in the HSLS:09 data. In particular, I regress observed college aid in HSLS:09 on X_{1j}^{aidc}, X_{2j}^{aidc} interacted with student test scores (s_i), and X_{3j}^{aidc} interacted with family income (inc_i). The college-level variables in X include several measures constructed from the net tuition data in IPEDS. Because the parameters of the aid function equation are computationally expensive to compute, I don't include every data point relevant to college aid; instead, I include only summary measures that are found to be empirically important in the auxiliary model. Specifically, by college, I summarize the aid data by taking the average net tuition paid by students above and below the middle income group. I also generate the gradient of net tuition in the three lowest and three highest income groups. The explanatory variables included in the regression are in Table 5. I also match the proportion of students in each of the four institutional aid categories in order to estimate the cutpoints in the institutional aid function. The estimates for the auxiliary institutional aid regression and the proportion of student in each aid category are shown in Table 5.

To summarize, the GMM objective function is formed using the following moments, where the # moments = #parameters:

$$E[\mathbf{Z_{ijl}}\xi_{jl}] = 0$$
$$E[\mathbf{X_{jl}^{c}}\xi_{jl}^{c}] = 0$$
$$E[\mathbf{b^{model}} - \mathbf{b^{aux}}] = 0,$$

where $\mathbf{b^{aux}}$ are the moments from the HSLS:09 data, and $\mathbf{b^{model}}$ are the model counterparts.

To use the first two moments above requires having the values of ξ_{jl} and ξ_{jl}^{c} as a function of the model parameters. This is a common situation when estimating models of demand from market shares. Analytic inversion of the market shares for the mean utility is often impossible or prohibitively difficult, as it is in my case (Berry, Levinsohn, and Pakes 1995). In my setting there is an additional twist relative to the traditional framework, as there are two state-by-college level data shares to match: enrollment shares and application penetration. Recall that ν_{jl}, the mean value of utility at the college-market level, is linear in certain parameters ($\theta^{l,\nu}$), data, and the unobserved college quality ξ_{jl}. The mean application cost, c_{jl} is a function of another set of linear parameters ($\theta^{l,c}$) and the college-state level unobservable in application cost. Both the mean utility and the mean application cost are functions of a common set of non-linear parameters, θ^{nl}.

Equations 10 and 12 cannot be inverted analytically for $\boldsymbol{\nu}$ and $\mathbf{c}$, but a fixed-point algorithm similar to

those used in Rust (1987) and Berry, Levinsohn, and Pakes (1995) allows numeric inversion, conditional on values for θ^{nl}. I stack the following two equations and iterate until convergence:

$$\boldsymbol{\nu}^{h+1} = \boldsymbol{\nu}^{h} + \log(\mathbf{E}) - \log(e(\boldsymbol{\nu}^{h}, \mathbf{c}^{h}, \theta^{nl}))$$

$$\mathbf{c}^{h+1} = \mathbf{c}^{h} - (\log(\mathbf{App}) - \log(app(\boldsymbol{\nu}^{h}, \mathbf{c}^{h}, \theta^{nl})),$$

where E and App are the enrollment shares and application penetration in the data, respectively, and $e(\boldsymbol{\nu}^h, \mathbf{c}^h)$ and $app(\boldsymbol{\nu}^h, \mathbf{c}^h)$ are the model-derived counterparts. Once the $\boldsymbol{\nu}$ and $\mathbf{c}$ are obtained, the linear utility and application cost parameters are found using IV-GMM, where the $\boldsymbol{\nu}$ and $\mathbf{c}$ are the dependent variables. The unobserved quality and unobserved component of application cost are constructed as the residuals.

Once I obtain the residuals, I construct the GMM objective function from the moments listed above. Using the strategy of Berry, Levinsohn, and Pakes (1995), I search for the parameters θ^{nl} that set this GMM objective function to zero. In every iteration of the search for θ^{nl}, I compute the residuals ξ_{jl} and ξ^{c}_{jl} using the fixed-point algorithm described above.

5.5 College Choice and Financial Aid Policy Parameters: Results

Table 6 shows the estimates for the institutional aid function. These can be used to find the college aid policy parameters, following the specification in Equation 15. Figure 4 illustrates the institutional aid policies by showing the expected aid amount by ACT score and income level for four example universities. Overall, the two illustrative private universities (Tulane and Loyola) give more aid than the example public universities (LSU and LA Tech). LSU and LA Tech give relatively more merit aid than the two private universities, in the sense that the aid gradient is very steep in the ACT score.

Table 7 shows the enrollment utility estimates. These are most easily interpreted through a few simulations for the demand-side alone. Table 8 shows the total applications and enrollment under two counterfactual simulations assuming there are no supply-side adjustments. In the first simulation, I decrease tuition by 1% at all colleges. Financial aid probabilities and amounts remain fixed, so all students are affected by this tuition decrease, even if they receive financial aid. Looking at Table 8, we see that a 1% decrease in tuition at all colleges is predicted to increase the number of applications sent by 10,021 (0.19%) and the number of students enrolled by 10,681, or 1.07% of baseline enrollment. This suggests that market demand

is modestly elastic with respect to prices. To help contextualize the estimates, the total high school graduate population is 3.4 million, so approximately 30% of all high school graduates enroll, as the baseline enrollment is nearly 1 million. Decreasing the admission thresholds by 1% at all colleges increases applications by 5,135 (0.10%), and enrollment by 6,324 (0.63%). Many universities already have very low admissions thresholds, which is why we might expect that the choice of college attendance vs. no college attendance is not very sensitive to the admissions process. However, enrollment at specific high-selectivity institutions may be much more sensitive to admissions thresholds.

The bottom five rows in Table 8 show the effects of the two simulated changes on enrollment by income level. The highest-income students (family incomes above $110k) are less price-sensitive, with an overall enrollment elasticity of 0.9%. There is a very strong relationship between income and the enrollment effect of an admissions threshold decrease. A 1% decrease in admissions thresholds across all colleges increases enrollment for the lowest income group by 0.76%, but only by 0.48% for the highest-income group. There are a few factors at play here: the utility estimates show that students with higher family incomes have a greater preference for college attendance, even apart from differences in price sensitivity (see the coefficient on income in Table 7), so higher-income students are more likely to apply to and attend a college than a lower-income student with the same ACT score. Also, ACT scores are correlated with income in the population, making higher-income students more likely to reach the admissions standards before the admission threshold change.

The two simulations shown so far are market-level changes, but I can also use the model to examine demand properties at the institution-level. I simulate a 1% decrease in tuition at LSU and Loyola (New Orleans), and find that a 1% decrease in tuition at LSU generates a 0.93% increase in enrollment at LSU, while at Loyola a 1% decrease in tuition would increase enrollment by 4%. Cross-price elasticities for these universities and their competitors are shown in Figure 5. Panel A shows the top 25 universities in terms of cross-price elasticity with respect to LSU. The largest cross-price elasticities (around -0.12) are for regional public universities in Louisiana (e.g., University of Louisiana in Lafayette, Nicholls State University, and University of New Orleans) and relatively non-selective private universities in the New Orleans area, such as Dillard University and Xavier University. Panel B shows the top 25 universities in terms of cross-price elasticities with respect to Loyola (New Orleans). Note the difference in the y-axis between LSU and Loyola: since Loyola's enrollment is relatively low and is spread out across the country (as shown in Figure 2), the cross-price elasticity between Loyola and any specific competitor university is smaller. However, there is

still a strong regional component to Loyola's enrollment, making moderately or less-selective institutions in Louisiana its strongest competitors, as measured by cross-price elasticity. Other universities that show up in the top 25 include universities in Florida, Georgia, and Mississippi.

Finally, we can use the demand-side estimates alone to simulate the effects of a $1,000 increase in financial aid for public universities in Louisiana, holding the supply-side fixed. The effect on enrollment is shown in Figure 6. The additional grant aid increases enrollment between 5 and 8 percent at Louisiana in-state public institutions. It has little effect on highly-selective private universities that compete in Louisiana, such as Vanderbilt and Rice. This is because Louisiana students make up a very small proportion of these universities' enrollees. Similarly, out-of-state public universities don't have a large direct demand-side effect. However, less-selective private universities lose enrollment to the public universities, up to 4% of current enrollment. The net effect, with demand-side responses alone, is to increase enrollment by 1,015 students, or 5.7%.

5.6 Supply-Side Estimation and Results

On the supply-side, the model features college-specific marginal cost and preferences over profit, admission rates, mean academic fit of the student body, and average family income of the student body, as shown in Equation 13. The first-order conditions of the college problem form a system of equations. For private colleges and public colleges that compete only in-state, the system has four equations corresponding to the choice of tuition, admission threshold, and the financial aid policy coefficients on income and admission test scores. For public colleges that compete both in-state and out-of-state, the first-order conditions are a system of five equations, with the additional equation coming from the choice of the out-of-state tuition level. Letting $\boldsymbol{\theta_j^{sup}}$ denote the vector of choice variables for the college, the system of first-order conditions can be written as:

$$(1-\lambda_{1j}-\lambda_{2j}-\lambda_{3j})\frac{1}{\pi_j}\frac{\boldsymbol{\partial \pi_j}}{\boldsymbol{\partial \theta_j^{sup}}}+\lambda_{1j}\frac{1}{\tilde{s}_j}\frac{\boldsymbol{\partial \tilde{s}_j}}{\boldsymbol{\partial \theta_j^{sup}}}+\lambda_{2j}\frac{1}{r_j}\frac{\boldsymbol{\partial r_j}}{\boldsymbol{\partial \theta_j^{sup}}}+\lambda_{3j}\mu_j^{inc}\frac{\boldsymbol{\partial \frac{1}{\mu_j^{inc}}}}{\boldsymbol{\partial \theta_j^{sup}}}=\mathbf{0} \qquad (16)$$

The last three terms are functions of college-specific preferences and objects that can be calculated directly from the demand system at current values of the college choices. In other words, I can compute derivatives in the last three terms numerically, and then treat these derivatives as data.

The first term—which involves the profit function and its derivatives—is a function of the parameters $\boldsymbol{\alpha}$ and heterogeneous unobservables in cost. Recall that the cost per-student is $\alpha_1 + \alpha_2 q_j + \alpha_3 X_j^s + \omega_j$, so that the marginal cost is $mc_j = \alpha_1 + 2\alpha_2 q_j + \alpha_3 X_j^s + \omega_j$. Given a value for α_2, I can write $\frac{1}{\pi_j}\frac{\boldsymbol{\partial \pi_j}}{\boldsymbol{\partial \theta_j^{sup}}}$ as a function of demand-side derivatives and the linear component of cost, including the cost heterogeneity ω_j. For public universities, this term is also a function of the heterogeneous cost differential for in-state students ($\alpha_4 + \omega_j^{is}$).

Thus, given the value of α_2, I can solve the system of first-order conditions for each university separately to get the values of λ_{1j}, λ_{2j}, and λ_{3j}, along with the heterogeneous marginal cost ($mc_j = \alpha_1 + 2\alpha_2 q_j + \alpha_3 X_j^s + \omega_j$), and the heterogeneous cost differential for in-state students ($\alpha_4 + \omega_j^{is}$). The system has exactly as many equations as unknowns.[27] Once these solutions are obtained, I estimate α_1 and $\boldsymbol{\alpha_3}$ using OLS, where the dependent variable is $mc_j - 2\alpha_2 q_j$. Then, I can generate the marginal cost residuals ω_j, conditional on α_2.

I estimate the value of α_2 using GMM. Letting Z_j^{sup} be the instrument for enrollment in the marginal cost function, the moment used to estimate α_2 is $E[Z_j^{sup}\omega_j] = 0$. The process of estimating α_2 is iterative: for each value of α_2, I find the solutions to the first-order conditions and generate the marginal cost residuals using the process described in the previous paragraph. Using those residuals, I find the value of the GMM objective function. I search over values of α_2 until the GMM function is minimized.

The results of this process are summarized in Table 9, which shows the mean value of the preference parameters, the cost differential for in-state students, and the coefficients in the marginal cost function. The first three rows show the mean preference parameters (exponents λ_1, λ_2, and λ_3 in the college objective function) across colleges for the measure of academic fit, admission rate, and the reciprocal of mean income. The estimates suggest that on average, colleges seek to enroll a student body that is relatively lower-income and higher academic fit than what profit maximization alone would imply. Figure 7 shows the distribution of the preference parameters, demonstrating that there is substantial heterogeneity across colleges.

Given recent trends in admission rates at top colleges as documented by Blair and Smetters (2021), is perhaps surprising is that universities on average have a preference for a *higher* admission rate. My estimates are conditional on student academic fit; that is, provided a college could maintain the same average student academic performance, my estimates suggest that the college would like to admit a higher proportion of applicants. What drives this finding in the data? The college preference for academic fit is primarily

[27] I solve the first-order conditions using numerical methods. I also confirm numerically that the second-order conditions are met.

driven by the degree to which the college rewards ACT scores in the financial aid function: a high ACT score gradient in the financial aid function implies a higher preference for academic fit, all else equal. In contrast, admissions rates are largely a function of the admission threshold. Raising the threshold would have a significant positive impact on average student academic fit, but colleges choose not to. The effect on revenue is not enough to explain why colleges do not raise the admission threshold further, so the reason the admission threshold is not higher must be that universities on average do care about providing access.

The next section of Table 9 shows the marginal cost estimates. Private colleges are estimated to have a much higher marginal cost on average than public colleges, by approximately $12,910 per student. Colleges with a higher completion rate also have much higher marginal cost: a one percentage point increase in the completion rate is associated with a marginal cost increase of approximately $140. It is likely that private colleges and colleges with higher completion rates offer more support services and amenities to students, raising the cost per student. Other notable findings are that HBCUs have higher costs per student than other colleges, that colleges with higher room and board rates face higher costs (this is sensible as room and board rates are a function of real estate prices and cost of living in the area surrounding the college campus), and that the most-selective colleges as ranked by Barron's spend substantially more per-student than other colleges. Marginal cost is increasing the number of enrollees ($\alpha_2 = 0.013$), although the estimate is not statistically significant. The last row of the table shows that the average benefit to public universities for enrolling an in-state student, relative to an out-of-state student, is $12,186.

6 Counterfactual Financial Aid

In the last several decades, many state governments have created and expanded financial aid programs that subsidize students who attend in-state colleges. These programs typically favor public colleges, although some programs do offer aid for attending in-state private universities. Such policies affect student decisions, but they also affect college decisions, including colleges that aren't directly targeted by the policy. To illustrate the economic importance of the supply-side responses, I simulate the equilibrium responses to additional financial aid offered by an example state. Consistent with the examples throughout the paper, I choose Louisiana as the example state, although I could run this simulation for any state.[28]

I simulate the effects of an additional $1,000 grant to Louisiana students to attend in-state public univer-

[28] Equilibrium simulations are time-consuming and computationally expensive, which is why the counterfactuals in this draft are limited to a single policy change in a single state.

sities. As shown in Section 5.5, a comparable demand-side simulation (holding all college responses fixed) generated a 5.7% increase in enrollment. In this preliminary counterfactual, I allow public universities in Louisiana to respond by adjusting tuition and aid policies. Future work will expand and refine these counterfactuals using a utility function with additional dimensions of heterogeneity, allow private universities to respond to the policy, and allow universities to change admissions standards in the counterfactual. These are all feasible adjustments for future drafts.

When Louisiana public universities adjust their prices and aid policies in response to the counterfactual policy, I find that total Louisiana enrollment increases by 3.4%. This is a substantial reduction from the demand-side effect alone, which was to increase enrollment by 5.7%. Colleges benefit from the additional revenue generated by the demand-side subsidy, choosing to raise tuition net of institutional aid for especially for students with low test scores. When these students pay more, the college is able to increase merit-based aid to attract students with higher test scores. Note that revenue increases among students with low test scores will be driven by relatively more affluent students, as they are more likely to enroll conditional on test scores. Figure 8 illustrates the college responses, showing the expected tuition net of institutional grant aid before and after the policy for several public institutions in Louisiana, across a range of ACT scores and family income levels. For each university, there is a notable increase in merit aid. For example, at LSU, tuition minus expected institutional aid increases by over $500 for students with ACT scores of 15, while the increase for students with ACT scores of 30 is just over $200. These net tuition increases are offset by the $1,000 grant aid introduced by the state, so a prospective applicant to LSU with an ACT score of 15 would expect to pay $500 less to attend college after the subsidy ($800 less for a prospective applicant with an ACT score of 30). While less obvious, there is also a slight increase in the income gradient for each of the colleges, so students with higher family incomes should expect less benefit from the subsidy after accounting for college responses.

Table 10 shows the combined effect of the subsidy and public college responses on private college enrollment, holding fixed private college tuition levels, pricing policies, and admission standards. The colleges that have the most to lose from this policy are small, non-selective colleges in Louisiana, including Dillard University, Centenary College, Xavier University, and Loyola New Orleans. Other small private colleges located outside of the state, such as Spring Hill in AL and Millsaps in MS, still face non-negligible negative effects from the policy. Larger and more selective private colleges located further away, such as Vanderbilt, Baylor, or Rice, have very small negative enrollment effects. Overall, the model shows that the

effects of a state policy targeting public universities can be significant for private universities, especially those closest to the public universities in terms of location and selectivity.

7 Conclusions

The effects of state-level financial aid programs are complex, involving responses by students and colleges—including colleges that do not benefit directly from the policy. The experimental or quasi-experimental methods that are typically used to assess the effect of aid programs on students are not well-suited to examine policy effects at scale, taking into account the market structure of higher education. At the same time, standard market equilibrium models are not easily applied to this mixed market of public and non-profit private firms that care about characteristics of the students they enroll. This paper develops a equilibrium model that is appropriate for the unique aspects of this particular market. The model is distinct from previous models of higher education, as it takes into account the substantial heterogeneity in characteristics across colleges along with the geographic nature of the market.

The college choice framework involves three stages: application, admission, and enrollment. While the expressions derived from this model are not the same as those found in typical logit demand applications, I demonstrate that a combination of aggregate (state-level) data on enrollment and applications and a modified version of the fixed-point algorithm used in Berry, Levinsohn, and Pakes (1995) can be used to estimate the parameters of this model. I supplement the aggregate data with moments from an individual student survey to estimate college financial aid policies and allow for several dimensions of student-level heterogeneity in utility. I also use student-level survey data to estimate parameters governing the admission process.

On the supply-side, individual colleges set prices, aid policies, and admissions selectivity in a strategic environment. Universities (potentially) care about profit, but also their admission rate, a measure of student academic fit, and the mean income of the study body. I find heterogeneous preferences for each of these measures as solutions to a system of first-order conditions for each university's problem.

The estimated demand-side effects of financial aid are large: as an example, a $1,000 scholarship to students in Louisiana to attend an in-state public university increases enrollment by 5.7%. When public universities are allowed to respond to this policy by adjusting tuition and financial aid, they raise tuition and increase price discrimination, especially on the basis of merit. Taking these responses into account, the enrollment effect is just 3.4%. Private universities that compete closely with public universities lose as much

as 3% of their enrollment. This simulation suggests that subsidies in higher education can enable colleges to both raise tuition (as suggested by (Bennett 1987)), but also increase the degree of price discrimination in order to improve measures of student characteristics. This simulation suggests that increased government subsidies for higher education could be a potential explanation for the trend of increasing tuition paired with increased college financial aid in recent decades (Cook and Turner 2022).

Forthcoming work will expand and enrich this paper by adding additional dimensions of heterogeneity in utility. One aspect of heterogeneity that I plan to add would allow students from higher-income families to have a stronger preference for resources and amenities. In the current model, regional public universities can raise tuition on high-income low-ACT students in order to offer more aid to high-ACT students, but if high-income low-ACT students aren't likely to attend a public regional college because they prefer high-resource colleges, this may limit the public university's ability to move towards higher tuition and higher aid. A related concern is that high ACT students may have a stronger preference for high-resource universities, a pattern which is also not captured in the current estimates. Adding this dimension of heterogeneity will affect the degree to which lesser-quality universities can attract high-ACT students through financial aid. Future work will also extend the counterfactuals by allowing private universities and public universities outside Louisiana to respond to the subsidy.

Beyond these concerns that will be addressed in future iterations of this paper, there are limitations that are outside the scope of the current work and provide some direction for additional research. For example, this model does not allow for exit, although small private colleges could be significantly adversely affected by counterfactual policies that provide subsidies to public colleges. While this model predicts that they would stay open but adjust their prices, aid policies, and admission thresholds, some private colleges may in fact close with large changes in aid to public colleges. A model of college exit is an interesting direction for additional research. Another potential direction for future research is to incorporate endogenous quality differentiation. Finally, this model has assumed that there is no collusion or joint setting of endogenous variables. This may not hold in all cases, particularly for public universities run by governing boards or for universities that have strong incentives for collusion over tuition, admissions standards, and aid policies. There is a need for research examining the prevalence and possible effects of joint decision making among colleges.

References

Arcidiacono, Peter (2005). "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica* 73.5, pp. 1477–1524.

Bennett, William J. (1987). *Our Greedy Colleges*.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium". *Econometrica* 63.4, pp. 841–890.

Berry, Steven T (1994). "Estimating Discrete-Choice Models of Product Differentiation". *The RAND Journal of Economics* 25.2, pp. 242–262.

Blair, Peter Q. and Kent Smetters (2021). "Why Don't Elite Colleges Expand Supply?"

Bleemer, Zachary (2021). "Top Percent Policies and the Return to Postsecondary Selectivity".

Card, David and Alan B. Krueger (2005). "Would the elimination of affirmative action affect highly qualified minority applicants? Evidence from California and Texas". *Industrial and Labor Relations Review* 58.3, pp. 416–434.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2020). "Income Segregation and Intergenerational Mobility across Colleges in the United States". *Quarterly Journal of Economics* 135.3, pp. 1567–1633.

Cook, Emily E. and Sarah Turner (2022). "Progressivity of pricing at US public universities". *Economics of Education Review* 88.

Epple, Dennis, Richard Romano, Sinan Sarpça, and Holger Sieg (2017). "A general equilibrium analysis of state and private colleges and access to higher education in the U.S." *Journal of Public Economics* 155.September 2013, pp. 164–178.

Epple, Dennis, Richard Romano, Sinan Sarpça, Holger Sieg, and Melanie Zaber (2019). "Market power and price discrimination in the US market for higher education". *RAND Journal of Economics* 50.1, pp. 201–225.

Epple, Dennis, Richard Romano, and Holger Sieg (2006). "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education". *Econometrica* 74.4, pp. 885–928.

Fan, Ying (2013). "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market". *American Economic Review* 103.5, pp. 1598–1628.

Fillmore, Ian (May 2023). "Price Discrimination and Public Policy in the US College Market". *The Review of Economic Studies* 90 (3), pp. 1228–1264.

Fitzpatrick, Maria D and Damon Jones (2012). "High education, merit-based scholarships and post-baccalaureat migration".

Fu, C (2014). "Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market". *Journal of Political Economy* 122.2, pp. 225–281.

Fu, Chao, Junjie Guo, Adam J. Smith, and Alan Sorensen (2022). "Students' heterogeneous preferences and the uneven spatial distribution of colleges". *Journal of Monetary Economics* 129, pp. 49–64.

Gentzkow, Matthew (2007). "Valuing New Goods in a Model with Complementarity: Online Newspapers". *American Economic Review* 97.3, pp. 713–744.

Goldin, Claudia and Lawrence Katz (2008). *The Race between Education and Technology*. Belknap Press for Harvard University Press.

Gordon, Grey and Aaron Hedlund (2019a). "Accounting for the Rise in College Tuition". In: *Education, Skills, and Technical Change: Implications for Future U.S. GDP Growth*. Ed. by Charles Hultan and Valerie Ramey. University of Chicago Press, pp. 357–394.

— (2019b). "Accounting for Tuition Increases across U.S. Colleges".

Heckman, James J, Lance J Lochner, and Petra E Todd (2006). "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond". In: *Handbook of the Economics of Education*. Vol. 1.

Herbaut, Estelle and Koen Geven (2020). "What works to reduce inequalities in higher education? A systematic review of the (quasi-)experimental literature on outreach and financial aid". *Research in Social Stratification and Mobility* 65.November 2019, p. 100442.

Howell, Jessica S. (2010). "Assessing the Impact of Eliminating Affirmative Action in Higher Education". *Journal of Labor Economics* 28.1, pp. 113–166.

Hoxby, Caroline and Sarah Turner (2019). "The Right Way to Capture College "Opportunity"". *Education Next* 19.2.

Hoxby, Caroline M (1997). "How the Changing Market Structure of U.S. Higher Education Explains College Tuition".

Kapor, Adam (2020). "Distributional Effects of Race-Blind Affirmative Action". *Working Paper*.

Ma, Jennifer and Matea Pender (2022). *Trends in College Pricing and Student Aid 2022*. Tech. rep.

Page, Lindsay C. and Judith Scott-Clayton (2016). "Improving college access in the United States: Barriers and policy responses". *Economics of Education Review* 51, pp. 4–22.

Rothschild, Michael and Lawrence J White (1995). "The Analytics of the Pricing of Higher Education and Other Services in Which the Customers Are Inputs". *Journal of Polical Economy* 103.3, pp. 573–586.

Rust, John (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica* 55.5, pp. 999–1033.

Smith, Jonathan, Matea Pender, and Jessica Howell (2018). “Competition among Colleges for Students across the Nation”. *Southern Economic Journal* 84.3, pp. 849–878.

Table 1: College Characteristics

	Public		***Private***	
	Mean	SD	Mean	SD
Costs and Aid				
Tuition	8.37	2.45	32.47	8.78
Out-of-State Tuition	19.85	7.31	.	.
Room and Board	8.13	4.03	9.56	4.82
Avg. Grant All Sources, Income 0-30k and Fed Aid Recipient	11.48	3.93	31.01	13.03
Avg. Grant All Sources, Income 30-48k and Fed Aid Recipient	10.07	3.58	30.03	12.84
Avg. Grant All Sources, Income 48-75k and Fed Aid Recipient	7.03	3.14	27.24	12.02
Avg. Grant All Sources, Income 75-110k and Fed Aid Recipient	4.76	2.35	23.50	9.96
Avg. Grant All Sources, Income 110k+ and Fed Aid Recipient	4.24	2.16	18.69	6.17
Selectivity				
25th Percentile ACT	20.11	2.68	22.31	3.70
Barrons: Most Competitive	0.02	0.13	0.12	0.32
Barrons: Highly Competitive	0.06	0.24	0.11	0.32
Barrons: Very Competitive	0.18	0.38	0.40	0.49
Barrons: Competitive	0.50	0.50	0.30	0.46
Barrons: Less Competitive	0.18	0.38	0.05	0.22
Barrons: Non-Competitive	0.07	0.25	0.02	0.13
Other				
Completion Rate within 150% of normal time	0.56	0.15	0.68	0.15
Any Grad Program	0.97	0.17	0.82	0.38
Religious	0.00	0.00	0.57	0.49
Historically Black College or University	0.07	0.26	0.03	0.17
Carnegie Size and Setting: Small	0.10	0.30	0.56	0.50
Carnegie Size and Setting: Medium	0.44	0.50	0.34	0.47
Carnegie Size and Setting: Large	0.46	0.50	0.09	0.29
NCES Locale: City—Large	0.15	0.36	0.24	0.42
NCES Locale: City—Midsize	0.14	0.35	0.13	0.33
NCES Locale: City—Small	0.19	0.40	0.14	0.35
NCES Locale: Suburb	0.20	0.40	0.28	0.45
NCES Locale: Town or Rural	0.31	0.46	0.21	0.41
Observations	421		454	

NOTE: The table summarizes the characteristics of 4-year degree-granting private non-profit or public institutions in this analysis. All monetary values are in thousands of 2019 $.

Table 2: Federal Pell Grant Aid Regression

	coef	se
Income, $100,000s	1.215	2.450
Income Sq.	-12.359	2.960
Constant	4.543	0.464
var(e)	6.431	0.552
Obs.	800	

NOTE: Results from a Tobit regression of the Pell Grant amount on family income among students with family incomes below $75,000. The sample is limited to students with records in the National Student Loan Data System, so it includes all students who apply for federal aid. The Tobit regression has a lower limit of $0 and an upper limit of $5,645, corresponding the the Pell grant cap in the 2013-14 academic year. Income is measured in hundreds of thousands, and the Pell grant amount is measured in thousands.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-Up, and Student Records.

Table 3: State Grant Aid Functions

	(1) States with Large Merit Aid Programs		(2) CA		(3) All Other States	
	coef	se	coef	se	coef	se
Income, $100,000s	-0.788	0.250	-10.698	3.306	-5.241	0.463
Income Sq	0.093	0.032	1.379	0.706	0.481	0.068
ACT	0.649	0.251			0.607	0.384
ACT Sq	-0.010	0.005			-0.014	0.008
State Grant PS	3.900	0.230			7.158	0.950
Public	-1.404	0.346	0.103	4.516	-1.046	0.379
Constant	-8.442	3.013	6.293	4.660	-5.058	4.508
var(e)	4.355	0.380	67.763	19.146	15.678	1.359
Observations	370		80		830	

NOTE: Results from Tobit regressions of the student's state grant aid amount (in thousands) on their family income, ACT score, state grant aid program generosity as measured by the total state grant aid per high school graduate (State Grant PS), and an indicator for whether the student is attending a public or private university. The Tobit regressions each have a lower limit of $0. States are divided by type of grant aid program, as described in the text.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-Up, and Student Records.

Table 4: Admissions Parameters

	estimate	se
Variance of academic fit, σ_ζ^2	7.015	0.234
Admission Threshold: Constant	22.14	4.191
Admission Threshold: 25th Percentile ACT	-2.107	0.358
Admission Threshold: 25th Percentile ACT, Sq	0.078	0.008
Obs	9,790	

NOTE: Results obtained from a Probit regression of student admission outcomes on the student's ACT score and a quadratic in the 25th percentile of ACT scores at the college. The coefficient on the student ACT score is constrained to be 0, consistent with the empirical model. This allows for the identification of the variance in the Probit regression.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), High School Transcript and Second Follow-up. U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System 2013, Institutional Characteristics and Admissions Surveys.

Table 5: Target Moments from HSLS:09

	estimate	se
Panel A: Means		
Distance (100 mi)	1.223	0.048
Distance Sq.	7.157	0.981
Income (100,000s), Students at Private Universities	1.334	0.084
Income (100,000s), Students at Public Universities	1.039	0.020
Panel B: Institutional Aid Regression Coefficients from HSLS:09		
Private	6.336	0.908
Average Net Tuition, Income $<$ \$48k	-0.246	0.071
Average Net Tuition, Income $>$ \$75k	0.815	0.112
Net Tuition Gradient, Income $>$ \$48k	74.300	9.594
Net Tuition Gradient, Income $>$ \$48k, Squared	-0.004	0.001
Income	-0.192	0.413
Income * Average Net Tuition, Income $<$ \$48k	0.066	0.044
Income * Average Net Tuition, Income $>$ \$75k	-0.088	0.033
ACT	0.090	0.031
ACT*Average Net Tuition, Income $>$ \$75k	0.004	0.002
Panel C: Proportion of Students by Institutional Aid (Discretized)		
Institutional Aid = 0	0.366	0.013
Institutional Aid = \$1,280	0.137	0.009
Institutional Aid = \$5,430	0.252	0.011
Institutional Aid = \$21,230	0.246	0.011

NOTE: The table shows the moments that I match from student survey data. Panel A shows the mean distance to college and its square for all enrolled students. It also shows average family income at public and private universities. Panel B shows the results of a regression of college aid amounts on student and college data. The student characteristics are family income and ACT scores. The college data includes average net tuition paid by students within five different income categories. As explanatory variables in this regression, I use the average net tuition for the lowest and highest two income groups, and the gradient of net tuition across incomes in the three highest income groups. Panel C shows the proportion of students in each aid category. While aid is discretized by quartiles of the distribution, more than 25% have exactly 0 aid, which corresponds to the first category. Thus the first two categories do not represent 25% of the sample.

SOURCE: U.S. Department of Education, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Base Year, First Follow-Up, and Student Records.

Table 6: Institutional Aid Policy Estimates

	estimate	se
Panel A: Aid Policy Constant		
Private	0.430	0.893
Average Net Tuition, Income $<$ \$48k	-0.047	0.023
Average Net Tuition, Income $>$ \$75k	0.228	0.047
Net Tuition Gradient, Income $>$ \$48k	13.714	7.715
Net Tuition Gradient, Income $>$ \$48k, Squared	-0.001	0.004
Panel B: Aid Policy Coefficient on Income		
Constant	-1.581	3.530
Average Net Tuition, Income $<$ \$48k	-0.048	0.221
Average Net Tuition, Income $>$ \$75k	0.037	0.144
Panel C: Aid Policy Coefficient on ACT Score		
Constant	-2.298	0.551
Average Net Tuition, Income $>$ \$75k	-0.037	0.019
Panel D: Aid Policy Cut Points		
k1	2.467	1.687
k2	3.014	1.798
k3	4.231	1.548

NOTE: The table shows estimates of the institutional aid policy parameters. Panel A shows the parameters governing the constant in the aid policy for each university. Panel B shows the parameters governing the coefficients on income and ACT scores, respectively. Panel D shows the cut points that determine the aid category in the ordered logit aid function (see Equation 2).

Table 7: Utility Estimates

	coef	se
Panel A: Mean Utility Estimates		
Constant	-33.546	0.938
Religious	-0.922	0.158
Room and Board	-0.077	0.018
Public	-2.048	0.201
In-State	-2.531	0.143
Any Graduate Program	-0.554	0.213
Completion Rate (150 \% Normal Time)	3.679	0.650
HBCU	0.051	0.234
Barrons: Highly Competitive	-3.809	0.234
Barrons: Very Competitive	-4.434	0.245
Barrons: Competitive	-4.624	0.287
Barrons: Less Competitive	-4.246	0.370
Barrons: Non-Competitive	-3.845	0.400
Carnegie Size and Setting: Medium	-0.497	0.319
Carnegie Size and Setting: Large	0.234	0.464
NCES Locale: City—Midsize	2.005	0.288
NCES Locale: City—Small	0.387	0.211
NCES Locale: Suburb	0.900	0.216
NCES Locale: Town or Rural	0.488	0.212
Carnegie Classification: R1	-0.470	0.668
Carnegie Classification: R2	-2.707	0.640
Carnegie Classification: Doctoral/Research	-1.895	0.658
Carnegie Classification: Masters (larger programs)	-2.512	0.600
Carnegie Classification: Masters (medium programs)	-3.141	0.591
Carnegie Classification: Masters (smaller programs)	-2.091	0.615
Carnegie Classification: BA–Arts	-3.837	0.566
Carnegie Classification: BA–Diverse Fields	-2.578	0.580
Panel B: Student-College Interactions		
Net Tuition	-0.147	0.260
Distance	-1.448	0.417
Distance Sq	0.068	0.091
Income, 100,000s	0.441	0.589
Income x Price	0.009	0.012
Obs 875		

NOTE: Estimates of enrollment utility parameters. State fixed effects are included but not shown.

Table 8: Demand-Side Simulations

		1% Decrease in Tuition		1% Decrease in Admissions Thresholds	
Outcome Measure	Baseline Values	Change	% Change	Change	% Change
Apps	5,152,110	10,021	0.19%	5,135	0.10%
Enroll (Total)	999,863	10,681	1.07%	6,324	0.63%
Enroll: $0-$30k	197,388	2,125	1.08%	1,495	0.76%
Enroll: $30-$48k	209,947	2,378	1.13%	1,491	0.71%
Enroll: $48-$75k	150,015	1,818	1.21%	1,023	0.68%
Enroll: $75-$110k	125,843	1,501	1.19%	798	0.63%
Enroll: $110k+	316,672	2,858	0.90%	1,517	0.48%

NOTE: The table shows application and enrollment changes with two counterfactual simulations on the demand-side, holding the university choices fixed.

Table 9: Supply-Side Estimates

parameter	estimate	se
λ_1 (Mean Preference for Academic Fit)	0.597	0.007
λ_2 (Mean Preference for Admission Rate)	0.318	0.034
λ_3 (Mean Preference for Reciprocal of Mean Income)	0.023	0.001
Marginal Cost Parameters:		
α_1: Constant	6.933	7.285
α_2: Enrollment	0.0013	0.0020
α_3: Private	12.910	1.128
α_3: Completion Rate (150\% Normal Time)	14.074	3.929
α_3: Religious	-2.787	1.068
α_3: Room and Board	0.109	0.085
α_3: Any Graduate Program	1.016	1.509
α_3: HBCU	8.080	1.772
α_3: Barrons: Highly Competitive	-16.005	1.900
α_3: Barrons: Very Competitive	-17.376	1.861
α_3: Barrons: Competitive	-20.004	2.116
α_3: Barrons: Less Competitive	-20.031	2.513
α_3: Barrons: Non-Competitive	-20.740	2.936
α_3: NCES Locale: City—Midsize	3.006	1.265
α_3: NCES Locale: City—Small	2.322	1.223
α_3: NCES Locale: Suburb	2.554	1.122
α_3: NCES Locale: Town or Rural	3.482	1.174
α_3: Carnegie Classification: R1	-2.462	6.411
α_3: Carnegie Classification: R2	-3.447	6.367
α_3: Doctoral/Research Universities	-6.442	6.436
α_3: Carnegie Classification: Masters (larger programs)	-7.992	6.299
α_3: Carnegie Classification: Masters (medium programs)	-8.454	6.368
α_3: Carnegie Classification: Masters (smaller programs)	-4.882	6.455
α_3: Carnegie Classification: BA–Arts & Sciences	-2.490	6.239
α_3: Carnegie Classification: BA–Diverse Fields	-5.946	6.346
α_4 (In-state Student Differential)	12.186	0.446

NOTE: The table shows the supply-side cost and preference parameter estimates, as described in Section 5.6.

Table 10: Subsidy Effect on Private Universities After Public College Response

Institution	1st-Year Enrollment Before Subsidy	Percent Change in Enrollment
DILLARD UNIV	119	-3.05
CENTENARY COLL OF LA	88	-3.04
XAVIER UNIV	401	-2.40
LOYOLA UNIV NEW ORLEANS	345	-2.27
SPRING HILL COLL	199	-1.09
MILLSAPS COLL	122	-0.90
TULANE UNIV OF LA	1,399	-0.45
RHODES COLL	243	-0.45
MS COLL	444	-0.44
CLARK ATLANTA UNIV	474	-0.09
HOWARD UNIV	1,329	-0.08
RICE UNIV	691	-0.06
SOUTHERN METHODIST UNIV	1,010	-0.06
TX CHRISTIAN UNIV	1,467	-0.05
SPELMAN COLL	325	-0.05
BAYLOR UNIV	2,453	-0.04
VANDERBILT UNIV	1,160	-0.04
EMORY UNIV	1,208	-0.02

NOTE: The table shows the percentage change in private university enrollment when Louisiana students and public colleges respond to an additional $1,000 in grant aid for Louisiana students to attend in-state public colleges.

Figure 1: Model Timing

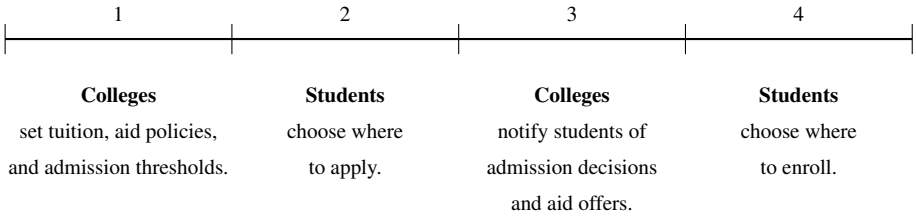
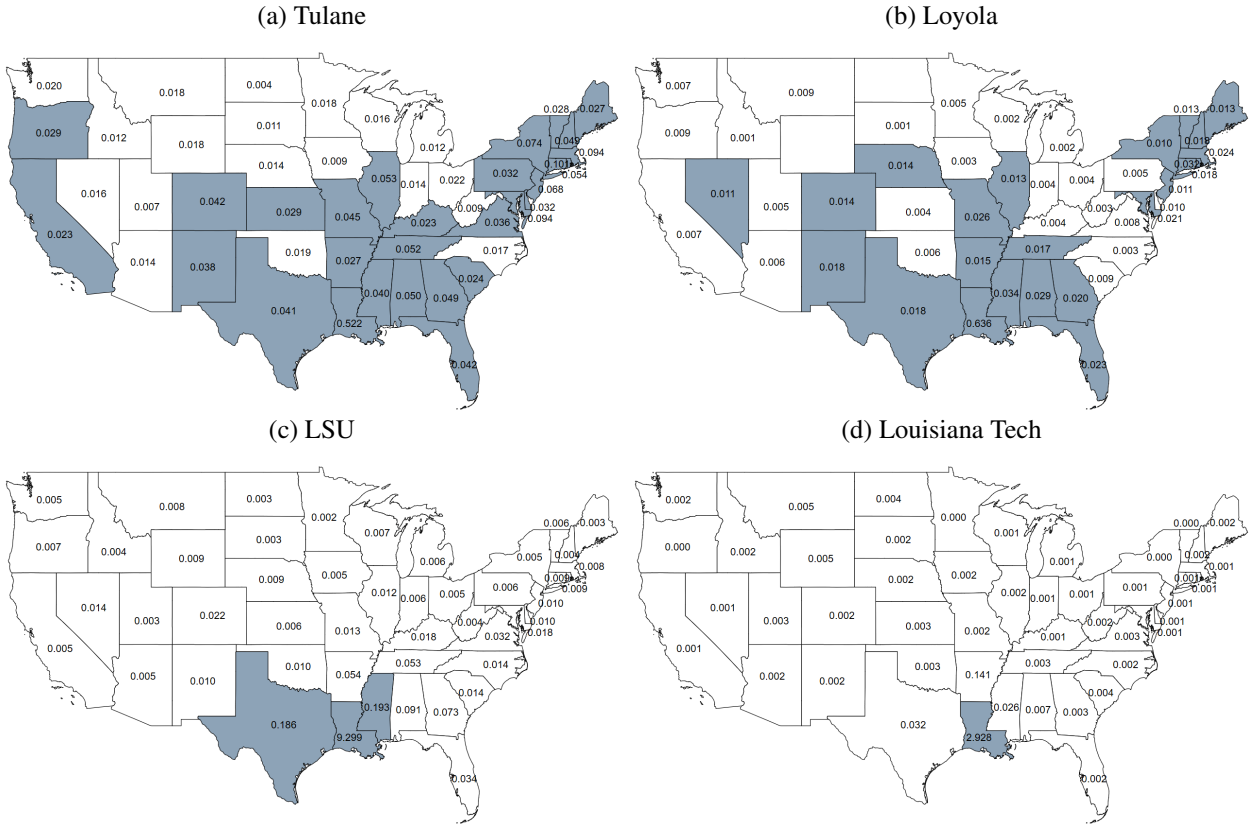
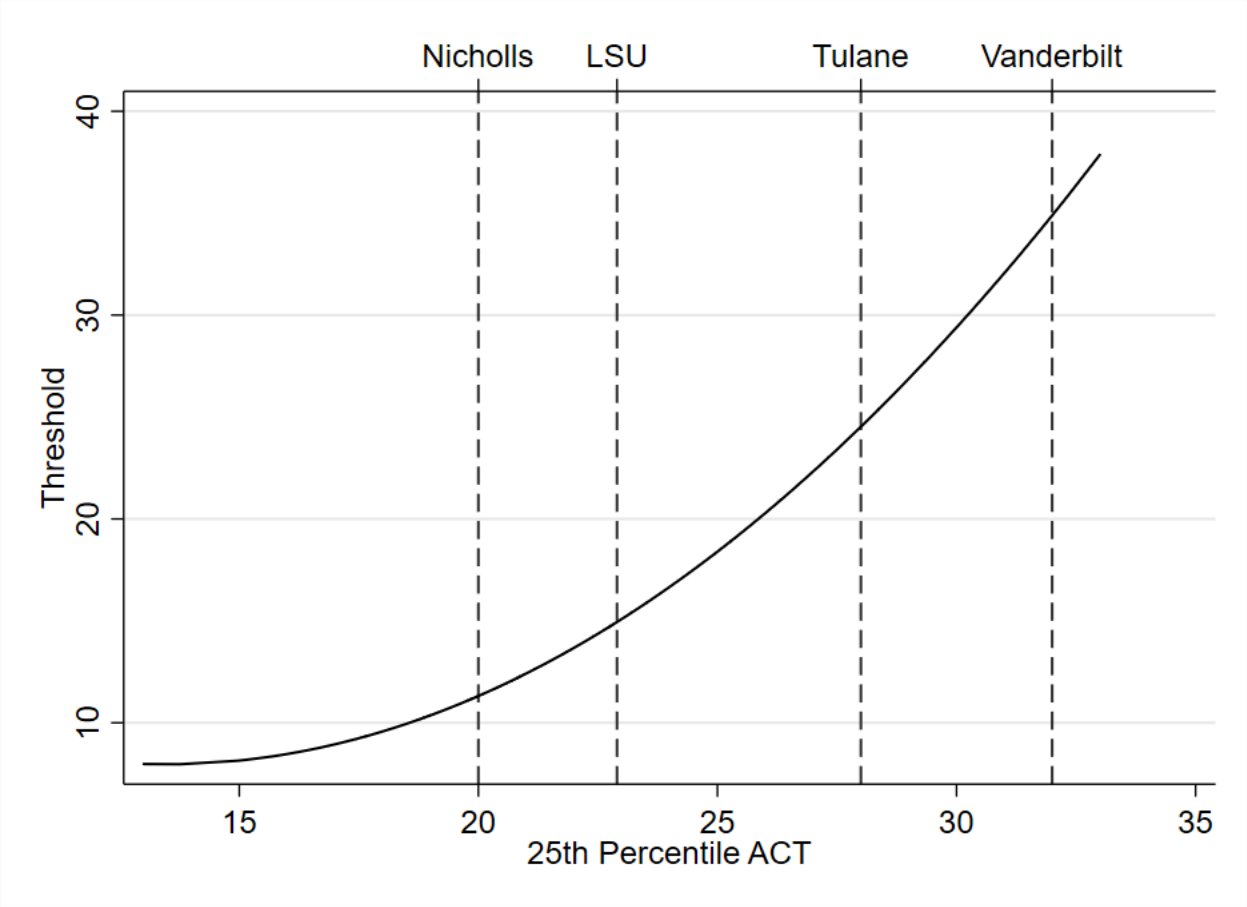


Figure 2: Market Definition



NOTE: The figure shows states by the percentage of high school graduates enrolled in the respective university on average between 2006 and 2012. For example, LSU enrolls 9.3% of Louisiana's high school graduating class on average during this period. Blue-shaded states are those in which the university competes, based on the method described in Section 2.3.

Figure 3: Admission Threshold Function and Example Universities

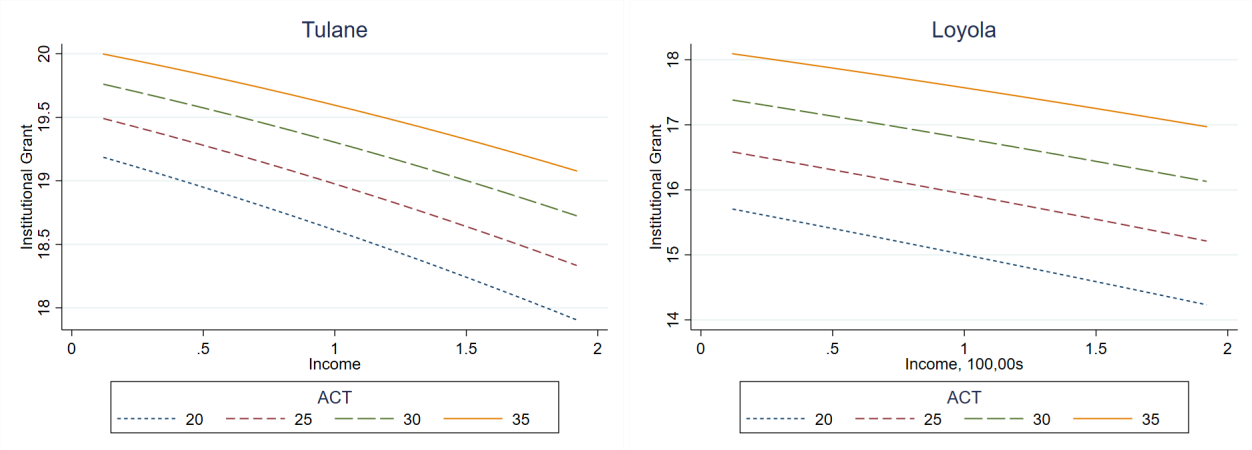


NOTE: The figure shows admission thresholds as estimated by the procedures described in Section 5.2. Example universities are shown for illustration. Vanderbilt and Tulane are selective private universities, while Louisiana State University (LSU) is the public flagship of Louisiana. Nicholls State is a regional public university in Louisiana.

Figure 4: Expected Institutional Aid by Income and ACT Score

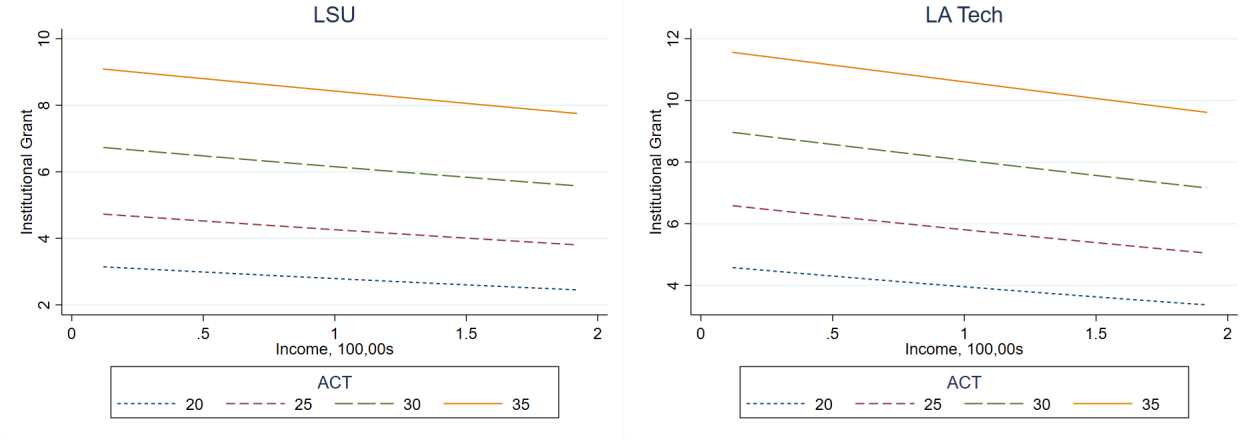
(a) Tulane

(b) Loyola



(c) LSU

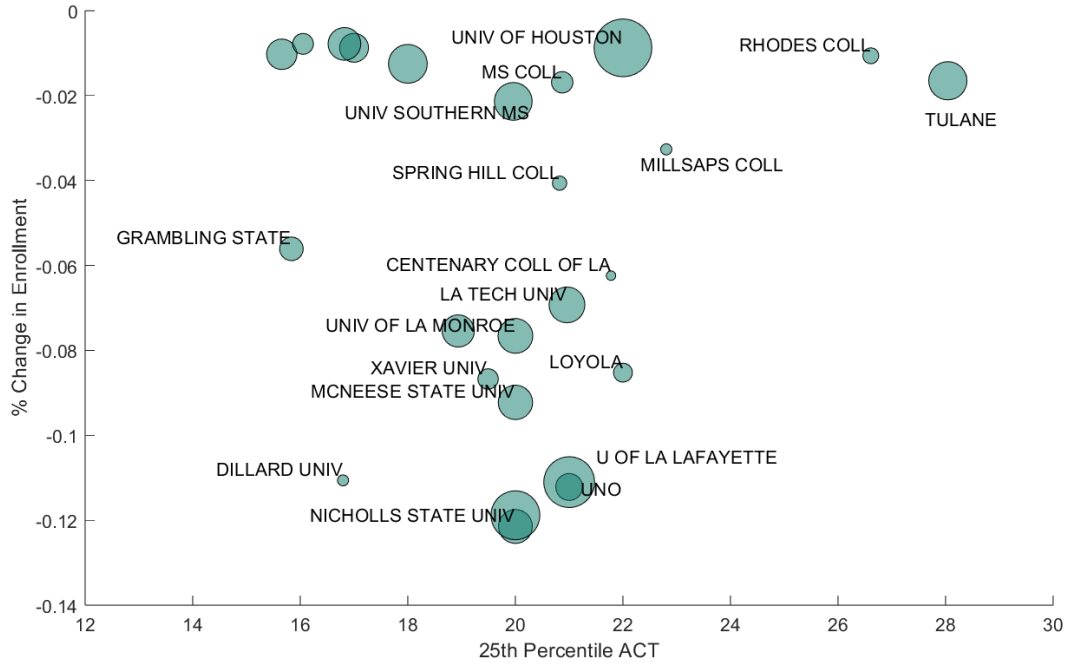
(d) Louisiana Tech



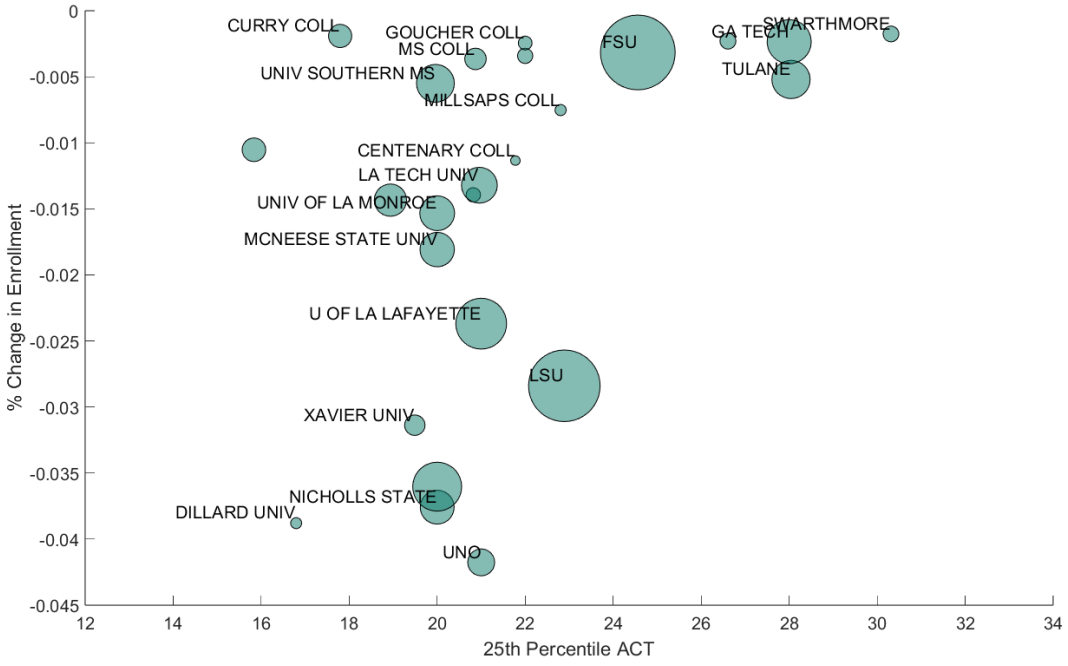
NOTE: The figure shows the expected aid amounts by income level and ACT score, for four selected universities.

Figure 5: Cross-price Elasticities for Example Universities

(a) LSU

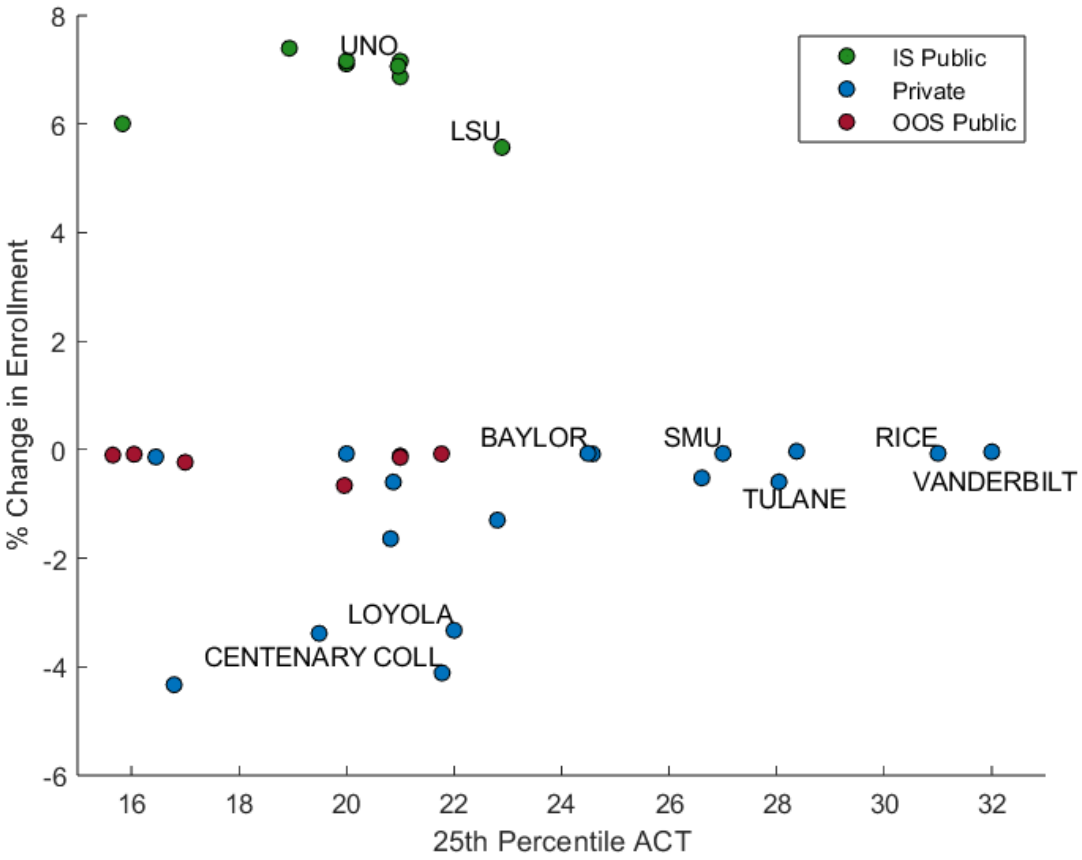


(b) Loyola



NOTE: Each subfigure shows the percentage change in enrollment at competitor universities in response to a 1% decrease in tuition for the university in the subfigure title, *ceteris paribus*. Thus, the percentage change in enrollment shown can be interpreted as cross-price elasticities. The universities shown are the top 25 competitors in terms cross-price elasticity with respect to the price of the university in the title.

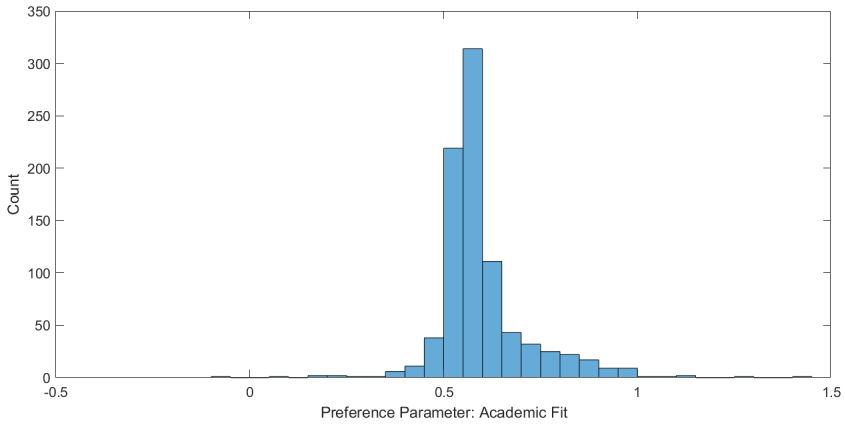
Figure 6: Demand-Side Responses to a $1,000 Scholarship for LA Public Universities



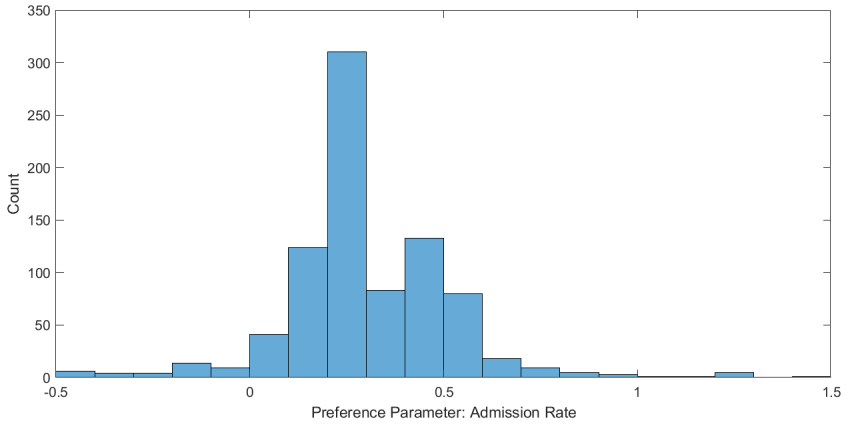
NOTE: The figure shows the percentage change in enrollment for universities that compete for Louisiana students when Louisiana reduces in-state tuition at all public universities by $1,000. These are demand-side effects only and assume that admissions standards remain fixed.

Figure 7: Supply-Side Results: Distribution of Heterogeneous Preferences

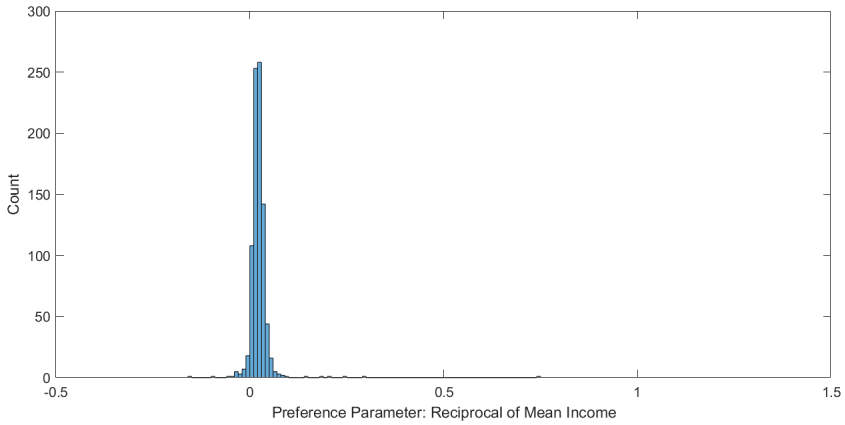
(a) Distribution of Preference for Academic Fit



(b) Distribution of Preference for Admission Rate



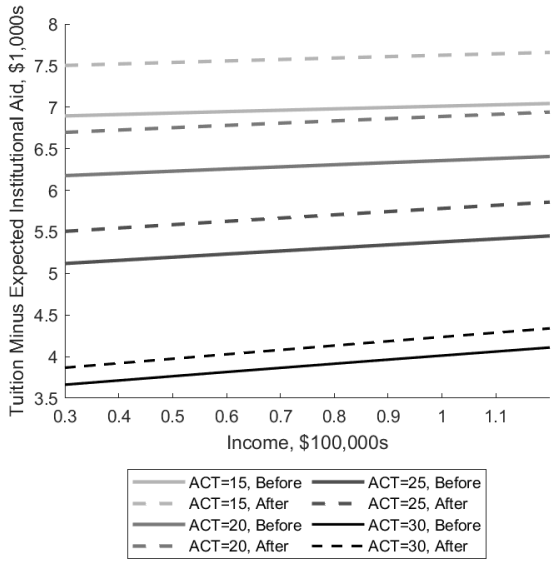
(c) Distribution of Preference for the Reciprocal of Mean Income



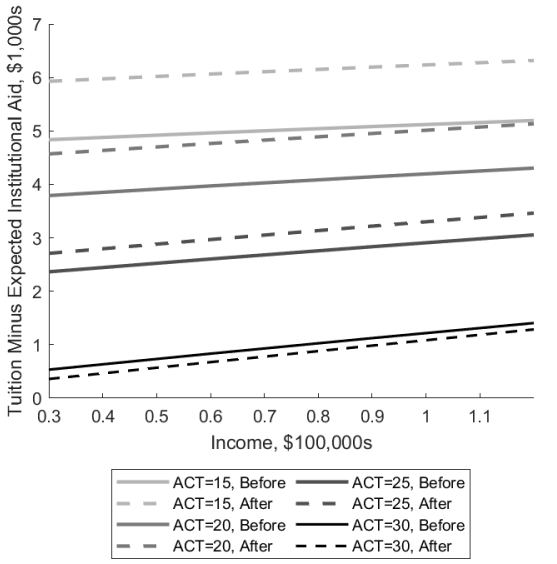
NOTE: The figure shows the distribution of preference parameters for mean academic fit, the admission rate, and the reciprocal of mean income across colleges. The results are described in the text in Section 5.6.

Figure 8: Counterfactual Results: Tuition Net of Institutional Grant Aid Before and After the Subsidy

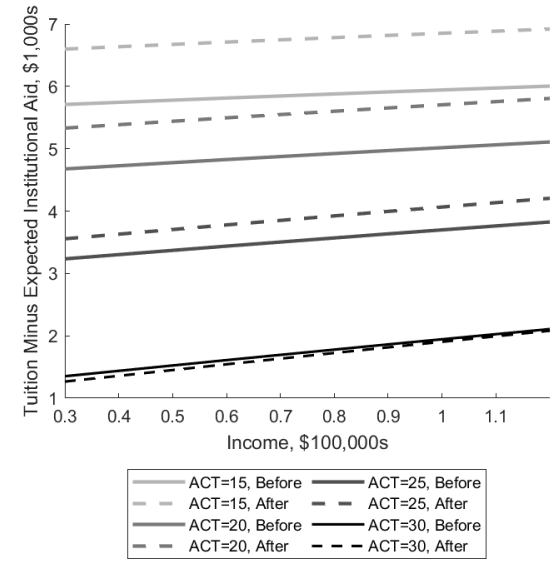
(a) LSU



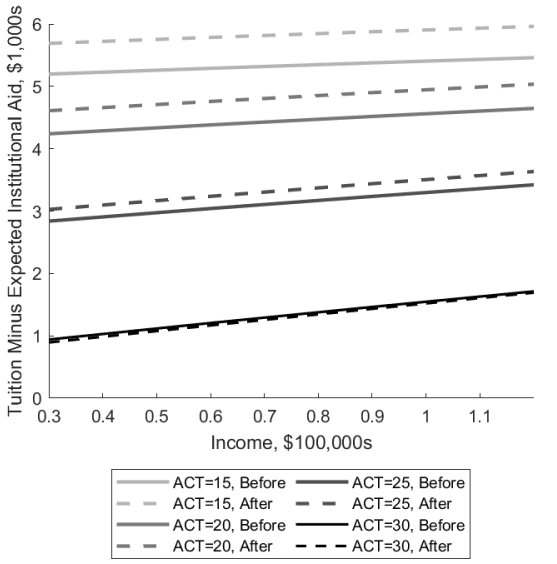
(b) University of New Orleans



(c) Louisiana Tech



(d) Nicholls State



NOTE: The figure shows the expected tuition net of institutional aid by institution, ACT score, and income level before and after the counterfactual subsidy described in the text.

Appendices

A Constructing a Representative Sample of Students

Estimation of the model requires a sample of students from each state that is representative in terms of income, admission test scores, and location. The Current Population Survey provides family income and location of respondents (up to the Public Use Microdata Area, or PUMA). I use these data to form the basis of my sample. Specifically, I use 2012-2014 CPS data for persons aged 17 or 18 who are in twelfth grade at the time of the CPS survey. I sample 1,000 of these individuals, weighting by the CPS person weights in order to ensure the sample is representative. Because PUMAs are sometimes very large, I assign individuals to Census tracts, based on the population distribution of Census tracts within each PUMA. I then calculate each students distance to each university based on the assigned Census tract. After this step, I have a representative sample of 1,000 students in each state along with their family income and distance to each university.

It remains to add ACT scores to my simulated sample. CPS does not provide college admission test scores, but I am able to find the joint distribution of admission test scores and income in the HSLS:09 data. Specifically, I run an ordered logit regression of ACT scores on a quadratic function of family income, estimating cut points for each ACT score. This regression then provides the probability distribution of ACT scores as a function of family income. I use this probability distribution to simulate ACT scores for my 1,000 simulated students in each state.

B Tables and Figures

Table B1: Data Example: Score Reports Sent by CT High School Graduates

State of Residence	University	University State	Score Reports Sent	% of HS Grads
CT	UNIV OF CONNECTICUT	CT	12,362	27.86
CT	CENTRAL CONNECTICUT STATE UNIV	CT	5,156	11.62
CT	EASTERN CONNECTICUT STATE UNIV	CT	4,315	9.73
CT	SOUTHERN CONNECTICUT STATE UNIV	CT	4,198	9.46
CT	QUINNIPIAC UNIV	CT	3,631	8.18
CT	UNIV OF RHODE ISLAND	RI	3,081	6.94
CT	NORTHEASTERN UNIV	MA	3,034	6.84
CT	WESTERN CONNECTICUT STATE UNIV	CT	2,889	6.51
CT	UNIV OF MASSACHUSETTS AMHERST	MA	2,542	5.73
CT	BOSTON UNIV	MA	2,455	5.53
CT	UNIV OF HARTFORD	CT	2,348	5.29
CT	UNIV OF VERMONT	VT	2,087	4.70
CT	UNIV OF NEW HAVEN	CT	1,928	4.35
CT	FORDHAM UNIV	NY	1,793	4.04
CT	BOSTON COLL	MA	1,518	3.42

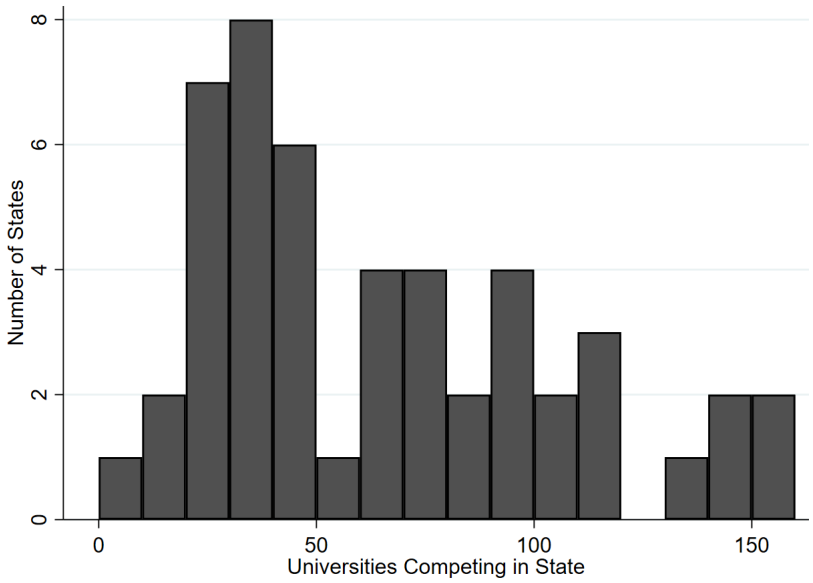
NOTE: This table shows the number of score reports sent and the percent of high school graduates sending reports to the top 15 colleges among high school graduates from Connecticut in 2013.

Table B2: Summary of Application and Enrollment Data

	Application Penetration	Enrollment Share
mean	0.02007	0.00451
min	0.00150	0.00001
p10	0.00214	0.00022
p25	0.00368	0.00040
p50	0.00784	0.00092
p75	0.01842	0.00276
p90	0.04629	0.00963
max	0.38405	0.16671

NOTE: This table shows the mean, min, max and percentiles of the enrollment and application variables used to estimate the college choice model. The application penetration for an institution is the proportion of high school graduates from a state market who sent a score report to the institution. The enrollment share is the proportion of high school graduates that enrolled at the institution.

Figure B1: Number of Colleges Competing in Each State



NOTE: This graph shows the distribution of the number of colleges competing in each state market.